(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 February 2002 (21.02.2002)

PCT

(10) International Publication Number
WO 02/14482 A2

(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number: PCT/GB01/03642

(22) International Filing Date: 14 August 2001 (14.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0019986.9 14 August 2000 (14.08.2000) GB

(71) Applicant *(for all designated States except US)*: **BIOTICA TECHNOLOGY LIMITED** [GB/GB]; 181A Huntingdon Road, Cambridge, Cambridgeshire CB3 0DJ (GB).

(72) Inventors; and

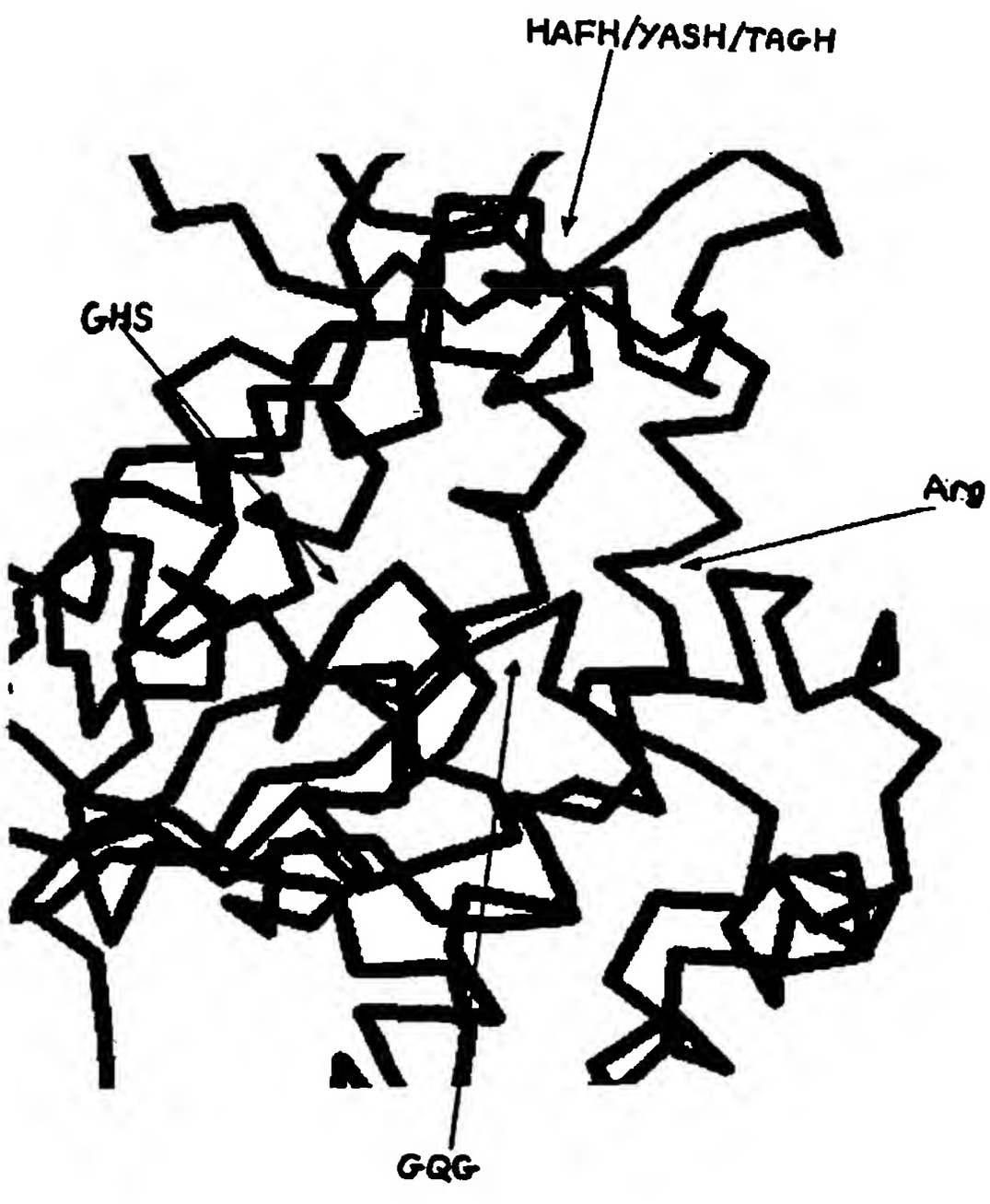
(75) Inventors/Applicants *(for US only)*: **PETKOVIC, Hrvoje** [SI/GB]; 89 Argyle Street, Cambridge, Cambridgeshire CB1 3LS (GB). **KENDREW, Steven, Gary** [GB/GB]; 6 Cowper Road, Cambridge, Cambridgeshire CB1 3SN (GB). **LEADLAY, Peter, Francis** [GB/GB]; 6 Westberry Court, Pinehurst, Grange Road, Cambridge, Cambridgeshire CB3 9BG (GB).

(74) Agents: **STUART, Ian** et al.; Mewburn Ellis, York House, 23 Kingsway, London, Greater London WC2B 6HP (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

[Continued on next page]

(54) Title: POLYKETIDES AND THEIR SYNTHESIS



(57) Abstract: Biosyntheses of compounds whereof at least portions are polyketides produced by means of polyketide synthase (PKS) enzyme complexes are carried out after specific alterations have been made within the acyltransferase (AT) domains of the PKSs. Particular motifs in or near the substrate binding pocket are disclosed, such that alterations therein affect substrate specificity.

WO 02/14482 A2

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Polyketides and Their Synthesis

Technical Field

The present invention relates to processes and materials (including enzyme systems, nucleic acids, vectors and cultures) which can be used to influence the selection of acylthioester units for the synthesis of polyketides, and to the resulting polyketides, which may be novel. It is particularly concerned with macrolides, polyethers or polyenes and their preparation making use of recombinant synthesis.

In preferred types of embodiment, polyketide biosynthetic genes or portions of them, which may be derived from different polyketide biosynthetic gene clusters, are manipulated to allow the production of specific polyketides, such as 12-, 14- and 16-membered macrolides, of predicted structure. The invention is particularly concerned with the modification of an Acyl CoA:ACP transferase (AT) function, generally by modifying genetic material encoding it in order to prepare polyketides with a predetermined ketide unit, e.g. incorporating (a) a malonate extender unit; or (b) a methylmalonate extender unit; or (c) an ethylmalonate extender unit; or (d) a further type of extender unit; or (e) an acetate and/or malonate starter unit; or (f) a

propionate and/or methylmalonate starter unit; or (g) a butyrate and/or ethylmalonate starter unit; or (h) a further type of starter unit. Of course the invention can be used to influence more than one ketide unit of a polyketide. The method enables one to minimise alteration to the protein structure of the polyketide synthase.

Polyketides are a large and structurally diverse class of natural products that includes many compounds possessing antibiotic or other pharmacological properties, such as erythromycin, tetracyclines, rapamycin, avermectin, monensin, epothilone and FK506. In particular, polyketides are abundantly produced by *Streptomyces* and related actinomycete bacteria. They are synthesised by the repeated stepwise condensation of acylthioesters in a manner analogous to that of fatty acid biosynthesis. The structural diversity found among natural polyketides arises in part from the selection of (usually) acetate (malonyl-CoA) or propionate (methylmalonyl-CoA) as "starter" or "extender" units (although one of a variety of other types of unit may occasionally be selected); as well as from the differing degree of processing of the β-keto group formed after each condensation. Examples of processing steps include reduction to β-hydroxyacyl-, reduction followed by

dehydration to 2-enoyl-, and complete reduction to the saturated acylthioester. The stereochemical outcome of these processing steps is also specified for each cycle of chain extension. Methylation at the α-carbon or β-hydroxy is also sometimes observed.

The biosynthesis of polyketides is performed by a group of chain-forming enzymes known as polyketide synthases. Two broad classes of polyketide synthase (PKS) have been described in actinomycetes. One class, named Type I PKSs, represented by the PKSs for the macrolides erythromycin, oleandomycin, avermectin, and rapamycin and by the PKS for the polyether monensin, consists of a different set or "module" of enzymes for each cycle of polyketide chain extension. For an example see Figure 1 (Cortés, J. *et al*. Nature (1990) 348:176-178; Donadio, S. et *al*. Science (1991) 2523:675-679; Swan, D.G. *et al*. Mol. Gen. Genet. (1994) 242:358-362; MacNeil, D. J. *et al*. Gene (1992) 115:119-125; Schwecke, T. *et al*. Proc. Natl. Acad. Sci. USA (1995) 92:7839-7843; also Patent application WO98/01546). The genes encoding numerous Type I PKSs have been sequenced and these sequences disclosed in publicly available DNA and protein sequence databases including Genbank, Swissprot and EMBL. For example, the sequences are available for the PKSs

governing the synthesis of erythromycin (Cortes, J. *et al*. Nature (1990) 348:176-178); accession number X62569, Donadio, S. *et al*. Science (1991) 252:675-679; accession number M63677); rapamycin (Schwecke, T. *et al*. Proc. Natl. Acad. Sci. (1995) 92:7839-7843; accession number X86780); rifamycin (August, P.*et al*. Chem. Biol. (1998) 5:69-79; accession number AF040570) and tylosin (Eli Lilly, accession number U78289), among many others.

The term "polyketide synthase" (PKS) as used herein refers to a complex of enzyme activities responsible for the biosynthesis of polyketides. These enzyme activities include β-ketoacyl ACP synthase (KS), acyltransferase (AT), acyl carrier protein (ACP), β-ketoreductase (KR), dehydratase (DH), enoylreductase (ER) and thioesterase (TE) but are not limited to these activities. Each of these activities lies on a separate protein or polypeptide fragment responsible for this activity. Such a fragment is termed a "domain". The terms "motif" or "signature sequence" used herein refer to a small stretch of amino acids (usually less than 10 amino acids) within a domain responsible (at least in part) for one aspect of the catalytic function, for example, choice of substrate.

The term "extension module" as used herein refers to the set of contiguous domains, from a β-ketoacyl-ACP synthase

("KS") domain to the next acyl carrier protein ("ACP") domain, which accomplishes one cycle of polyketide chain extension; this may or may not include domains responsible for the reductive processing of the polyketide chain. The term "loading module" is used to refer to any group of contiguous domains that accomplishes the loading of the starter unit onto the PKS and thus renders it available to the KS domain of a specific extension module.

Background Art

Several approaches to altering the nature of the polyketide product of a PKS by genetic engineering have been proposed: see particularly WO 93/13663 and WO 98/01571. The length of polyketide formed has been altered, in the case of erythromycin biosynthesis, by specific relocation using genetic engineering of the enzymatic domain of the erythromycin-producing PKS that contains the chain-releasing thioesterase/cyclase activity (Cortés, J.*et al*. Science (1995) 268:1487-1489; Kao, C.M. *et al*. J. Am. Chem. Soc. (1995) 117:9105-9106).

In-frame deletion of the DNA encoding part of the ketoreductase domain in module 5 of the erythromycin-producing PKS (also known as 6-deoxyerythronolide B synthase, DEBS) has been shown to lead to the formation

of erythromycin analogues 5,6-dideoxy-3-α-mycarosyl-5-oxoerythronolide B, 5,6-dideoxy-5-oxoerythronolide B and 5,6-dideoxy, 6 β-epoxy-5-oxoerythronolide B (Donadio, S. *et al*. Science (1991) 252:675-679). Likewise, alteration of active site residues in the enoylreductase domain of module 4 in DEBS, by genetic engineering of the corresponding PKS-encoding DNA and its introduction into *Saccharopolyspora erythraea*, led to the production of 6,7-anhydroerythromycin C (Donadio, S. *et al*. Proc Natl. Acad. Sci. USA (1993) 90:7119-7123).

Patent application WO 00/01827 describes further methods of manipulating a PKS to change the oxidation state of the β-carbon. Substituting the reductive domain of module 2 of the erythromycin-producing PKS with domains derived from rapamycin PKS modules 10 and 13 led to the formation of C10-C11 olefin-erythromycin A and C10-C11 dihydroerythromycin A respectively.

The second class of PKS, named Type II PKSs, is represented by the synthases for aromatic compounds. Type II PKSs contain only a single set of enzymatic activities for chain extension and these are re-used as appropriate in successive cycles (Bibb, M. J. *et al*. EMBO J. (1989) 8:2727-2736; Sherman, D. H. *et al*. EMBO J. (1989) 8:2717-2725; Fernandez-Moreno, M.A. *et al*. J.

Biol. Chem. (1992) 267:19278-19290). The "extender" units for the Type II PKSs are usually acetate (malonyl-CoA) units, and the presence of specific cyclases dictates the preferred pathway for cyclisation of the completed chain into an aromatic product (Hutchinson, C. R. and Fujii, I. Annu. Rev. Microbiol. (1995) 49:201-238). Hybrid polyketides have been obtained by the introduction of cloned Type II PKS gene-containing DNA into another strain containing a different Type II PKS gene cluster, for example by introduction of DNA derived from the gene cluster for actinorhodin, a blue-pigmented polyketide from *Streptomyces coelicolor*, into an anthraquinone polyketide-producing strain of *Streptomyces galileus* (Bartel, P. L. *et al*. J. Bacteriol. (1990) 172:4816-4826). Occasionally, unusual starter units are incorporated by Type II PKS, particularly in the biosynthesis of oxytetracycline, frenolicin and daunorubicin and in these cases a separate AT is used to transfer the starter unit to the PKS.

Fungal PKSs such as the 6-methylsalicylic acid or lovastatin PKS typically consist of a single multi-domain polypeptide which include most of the activities required for the synthesis of the polyketide portion of these molecules (Hutchinson C.R. and Fujii I. Annu. Rev. Microbiol. (1995) 49:201-238). Type II Fungal PKSs are

also known.

A number of mixed systems comprising polyketide synthase and nonribosomal peptide synthase modules have been identified including the epothilone and bleomycin biosynthetic clusters.

Although large numbers of therapeutically important polyketides have been identified, there remains a need to obtain novel polyketides that have enhanced properties or possess completely novel bioactivity. The complex polyketides produced by Type I PKSs are particularly valuable, in that they include compounds with known utility as anthelminthics, insecticides, anticancer, immunosuppressants, antifungal or antibacterial agents. Because of their structural complexity, such novel polyketides are not readily obtainable by total chemical synthesis, or by chemical modifications of known polyketides. Particular changes that are desired are changes to the carbon skeleton by altering the nature of the starter and/or extender unit(s) incorporated, changes to the oxidation level of the β-keto carbon and therefore the pattern of oxygen substituents by altering the series of reductive steps that occur after chain extension and changes to the post PKS "tailoring" steps which generally comprise hydroxylation, methylation or glycosylation of the polyketide molecule.

There is also a need to develop reliable and specific ways of deploying individual modules in practice so that all, or a large fraction, of hybrid PKS genes that are constructed, are viable and produce the desired polyketide product. Various strategies have been described to produce these hybrid PKSs particularly utilising recombinant DNA technology and denovo biosynthesis. There is a particular need to develop methods of manipulating these PKS in a manner that minimises the alteration to the PKS protein structure. Existing methods of achieving these manipulations sometimes produce hybrid PKS multienzymes which give the desired product at only 1% or less of the rate that the unmodified PKS produces product.

WO 93/13663 and WO 98/01571 describe novel methods of engineering PKSs. A well-established method of altering the nature of the extender unit used at any position in the polyketide molecule, particularly malonyl-, methylmalonyl- or ethylmalonyl-CoA is by domain substitution. For example, WO98/01546 and US patent 6,063,561 disclose methods of accomplishing this modification to form modified erythromycins. Novel polyketide molecules, in this case particularly novel erythromycins, are produced by the replacement of an entire AT domain-encoding DNA fragment on the

Saccharopolyspora erythraea chromosome with an equivalent heterologous AT domain-encoding fragment from another PKS cluster. It is well known to those skilled in the art that selection of the exact DNA/protein splice sites into which to insert the heterologous domain requires detailed analysis of the corresponding DNA and protein sequences. Different researchers choose to use splice sites at conserved, semi-conserved or non-conserved regions of the protein, or at sites either within or at the boundaries of the AT domains. A further drawback of this technique is that it is hard to predict whether a particular heterologous domain will work in any given context. A domain that works successfully in one module may not work at all in an adjoining module or may produce polyketides at a vastly reduced yield. Oliynyk, M. *et al*. (Chem. Biol. (1996) 3:833-839) and Ruan *et al*. (J. Bact. (1997) 179:6416-6425) have published studies that exchange a methylmalonyl-CoA specific AT domain for malonyl-CoA specific AT domains in modules of the erythromycin PKS. Products were observed only for changes in modules 1 and 2, with module 2 at a vastly lowered yield. Stassi *et al*. (Proc. Natl. Acad. Sci. (1998) 95:7305-9) exchange the methylmalonyl-CoA specific AT of module 4 of the erythromycin PKS for an ethylmalonyl-CoA specific AT and

again product yield was low even after the addition of the crotonyl-CoA reductase gene thought to increase the supply of the required ethylmalonyl-CoA precursor. A possible reason for the limiting yields is the structural or mechanistic non-compatibility of a heterologous AT domain with the adjoining KS and ACP domains with which it must interact properly for efficient polyketide chain synthesis. Consequently, it is often necessary to try multiple domain swaps to achieve a novel polyketide-producing strain that displays adequate efficiency - a process made particularly arduous when these changes must be made by gene replacement on the chromosome through a two step double integration process. The introduction of splice sites at the DNA level is time consuming and technically challenging, requiring careful analysis to ensure the PKS protein coding reading frame is not disrupted. The introduction of restriction enzyme sites often requires changes at the amino acid level which lead to further PKS protein structure disruption and consequent loss of catalytic efficiency.

A method that could utilise the numerous techniques available for site directed mutagenesis to influence the AT substrate specificity with minimal disruption to the protein tertiary structure would be a valuable addition to the current techniques.

Changes to an active site have been shown to alter substrate specificity in other systems. For example, in an early study, Scrutton *et al*. (Nature (1990) 343:38-43) used site directed mutagenesis to switch the coenzyme substrate specificity of a glutathione reductase. Identifying and changing a 'fingerprint' structural motif in the NADP+ binding domain they could convert the enzyme into one displaying a marked preference for NAD+. The techniques of directed evolution have been used to improve/change enzyme catalytic function. Of many examples in the literature, Zhang *et al*. (PNAS (1997) 94:4504-4509) illustrate the conversion of a galactosidase to a fucosidase by these techniques. The resulting protein bears 6 mutations, of which 3 lie in, or in close proximity to the active site.

Minor but directed changes to a PKS domain can make significant changes to its catalytic function. Patent application WO 00/00500 teaches that an extender ketosynthase domain is converted to a decarboxylating (and hence loading) ketosynthase domain by site directed mutagenesis at the active site. US Patent numbers 6,004,787 and 6,066,721 and Jacobsen *et al*. Science (1997)277:367-369 describe the deletion of residues in the KS1 active site to inactivate this activity to allow the production of novel polyketides by feeding of

synthetic precursors to the modified PKS.

Several studies have attempted to correlate the primary amino acid sequence of the AT to determine amino acids directly involved with the recognition of the appropriate substrate, and particularly the nature of the substrate side chain (i.e. the malonyl portion of the acyl-CoA thioester). Studies by Haydock *et al.* (FEBS Lett. (1995) 374:246-248) correlated the substrate specificity of malonyl- or methylmalonyl-CoA specific AT with a motif 11 amino acids upstream of the known active site. Comparisons between this motif and the protein structure of a known acyltransferase from *E. coli* fatty acid synthase allowed the authors to assess the proximity of the motif residues to the active site (and hence its ability to select the substrate). The authors acknowledged that "*this divergent region thus identified lies near the acyltransferase active site though not close enough to make direct contact with the substrate*".

Other studies (Katz, L. Chem Rev. (1997) 97:2557-2575, Tang, L. *et al.*, Gene (1998) 216:255-265) have correlated additional residues with a specific extender unit using these residues as a tool to predict the AT substrate specificity from a protein sequence derived from polyketide gene cluster sequencing projects. It has

remained unclear which residues have mechanistic importance. In only one case have regions within the PKS AT domain been exchanged in an attempt to swap AT specificity; patent application WO 00/01838 and Lau *et al*. Biochemistry (1999) 38:1643-51) implicated a 'hypervariable region' at the C-terminus of the AT domain in the selection of extender unit. These workers interchanged this 25-30 amino acid stretch and showed that this change was sufficient to alter the substrate specificity of the AT, concluding *"a short (23-35 amino acid) C-terminal segment present in all AT domains is the principal determinant of their substrate specificity. Interestingly its length and amino acid sequence vary considerably among the known AT domains. We therefore suggest that the choice of extender units by the PKS modules is influenced by a "hypervariable region", which could be manipulated via combinatorial mutagenesis to generate novel AT domains possessing relaxed or altered substrate specificity"*. Surprisingly, our structure molecular modelling studies indicate this region lies at a surface accessible region away from the active site and hence is unlikely to directly interact with (and hence directly select) the malonyl portion or the substrate

used. The effect on substrate specificity is therefore likely to be imprecise and due to more indirect effects via, for example, disruption of tertiary structure.

Disclosure of Invention

According to a first aspect of the present invention there is provided a method of synthesising a compound whereof at least a portion is the product of a polyketide synthase (PKS) enzyme complex or is derived from such a product, said PKS enzyme complex including at least one acyltransferase (AT) domain. The method includes a step of providing said PKS enzyme complex in which said AT domain has been altered to change selectively a minor proportion of amino acid residues. The altered residue(s) may comprise one or more motifs which are present in the active site pocket of the AT domain and which influence the substrate specificity of the AT domain, the alteration affecting the substrate specificity; and/or one or more residues of a motif which influences the substrate specificity of the AT domain and which comprises a four-residue sequence corresponding to the YASH motif of the AT domain of the first module of DEBS, the alteration affecting the substrate specificity.

Synthesis is then effected by means of said PKS enzyme complex to produce a compound or mixture of compounds

different from what could have been produced by means of a PKS enzyme in which said AT domain had not been altered.

The PKS enzyme complex may be at least part of a modular type I PKS enzyme complex, or it may be derived from a type II PKS system, a fungal PKS system or a hybrid system comprising PKS and nonribosomal peptide synthase modules.

The present invention teaches that by altering a few amino acid residues in the AT domain and particularly residues close to the AT active site comprising one or more residues of a short signature "motif" within the AT domain it is possible to influence the acylthioester selected by that AT domain. Novel polyketides can be made by a modified PKS on which the signature motif on one or more modules is altered, e.g. being replaced with one associated with a different specificity for malonyl substrate. Furthermore, the invention provides a method of reducing the proportion of mixed polyketide products that are occasionally found in natural systems due to non-specific incorporation of the incorrect extender units. Conversely, the invention provides a method of giving a mixed population of polyketide products thus increasing the diversity of polyketides produced by a PKS.

The invention allows the preparation of a modified PKS by substitution of an existing amino acid residue motif in the AT that specifies incorporation of one of the common extender acylthioesters with another motif found in another AT specifying an alternative acylthioester. This alters the substrate specificity of the polyketide synthase when it is expressed in a polyketide-producing organism.

The DNA sequences have been disclosed for numerous Type I PKS gene clusters. Comprehensive sequence analysis of AT domains derived from Type I PKS modules responsible for the formation of macrolides, particularly erythromycin, rapamycin, avermectin, rifamycin, FK506, epothilone, tylosin, and niddamycin, ionophore polyethers, particularly monensin, and polyenes, particularly nystatin, allowed us to identify amino acids that are characteristic of AT domains.

Figure 2 shows the sequence comparison of these AT domains. This sequence comparison has been generated in a generally conventional way, employing a computer using a procedure that creates a multiple sequence alignment from a group of related sequences. We used a program called PileUp (Wisconsin Package, Genetics Computer Group (GCG), Madison, WI,USA), which creates a multiple sequence alignment using simplification of the

progressive alignment method of Feng and Doolittle (Journal of Molecular Evolution 25; 351-360 (1987)). The method used is similar to the method described by Higgins and Sharp (CABIOS 5; 151-153 (1989)). The program executes a series of progressive, pairwise alignments that allows a large number of sequences to be compared together to form a final alignment throughout all the sequences. Gaps can be inserted throughout individual sequences to allow alignment of regions of strong similarity. This is often required as strongly conserved regions are often separated by more variable regions, both in terms of numbers of amino acids and type of amino acids. Different programs use different mathematical algorithms to make these comparisons, resulting in alignments that differ in minor ways. However, it can be expected that regions of strong homology would still align whatever alignment program is utilised. The particular motifs that are discussed are marked.

These motifs include the conserved GQG motif that is close to the start of the domain, the GHS motif that contains the active site serine that covalently binds the acyl chain prior to transfer to the ACP, and a LPTY motif that is close to the end of the domain. Other residues common to all ATs including an arginine, believed to stabilise the carboxylate group of the acylthioester.

Further detailed sequence analysis allowed us to identify amino acid residues that differed between ATs responsible for the incorporation of malonyl-, methylmalonyl- and ethylmalonyl-CoA. Some of these amino acids or motifs had been previously identified during the sequence analysis of the clusters as previously discussed. While these motifs could predict whether a malonyl-/methylmalonyl-CoA might be used they generally fail to show a difference between methylmalonyl- vs ethylmalonyl-CoA or the other larger extender unit commonly used. We viewed this as an important requirement for identification of the most important and key residues involved in substrate recognition and consequently residues most suitable for alteration. Closer analysis identified a string of four residues (location identified clearly in Figure 2) of which two residues are virtually invariant throughout all ATs, and two residues differ consistently depending on the extender unit. Particularly, in the vast majority of ATs responsible for recognition of malonyl-CoA the sequence of residues HAFH was identified. In the majority of ATs responsible for recognition of methylmalonyl-CoA the equivalent segment was substituted by residues YASH. In ATs responsible for ethylmalonyl-CoA or other similar sized CoA unit incorporation the overall motif was different, less

conserved but generally displayed the sequence XAGH (where X is most frequently but not limited to F, T, V or H). We typically use the terms HAFH, YASH and TAGH to describe these motifs with respect to malonyl-CoA, methylmalonyl-CoA and ethylmalonyl/further CoA specificity but use these terms herein to allow substitutions in the motif, particularly at residue 1 as described. Potential substitutions and the exact location of the motif will be clear to those skilled in the art by inspection of Figure 2 or similar sequence analysis.

There are three possible methods to locate the position of the motif within an AT sequence that does not appear in Figure 2. It is likely a combination of the methods will be used.

I) By simple visual inspection and comparison of the sequence to identify the motifs HAFH, YASH or TAGH. Since substitutions of residue one are often encountered a useful procedure is to look for an alanine (A) separated by one amino acid (usually F, S or G) from a histidine (H).

II) By counting amino acids from the active site serine. The start of the motif is typically (but should not be limited to) between 90 and

100 amino acids downstream of the GHS active site motif.

III) By computer generated multiple alignment that allows the new sequence to be directly compared to the sequences and motifs we have annotated in Figure 2 or to other ATs.

It is preferable to use the third method as this allows the motif to be identified unequivocally when there are substitutions within the motif. This is particularly necessary in some of the more unusual types of AT in which one of the residues can be substituted by proline (P). The third method will also identify the motif when the number of residues between the motif and the AT active site serine differs significantly from the norm. The third method will also better identify the motif when the same or similar string of amino acids occurs elsewhere in the domain.

A particular feature of these motif residues is the relationship of the size of the third residue compared to the substrate selected. Hence, when malonyl-CoA is required the third residue is large (phenylalanine), when methylmalonyl-CoA is required this residue is intermediate (serine), and when ethylmalonyl-CoA is required this residue is small (glycine). The inverse relationship between substrate side chain size and this

third residue is particularly noteworthy. Interestingly, this relationship applies also when considering the incorporation of the more unusual extender units such as methoxymalonyl-CoA, required for some cycles of chain extension during production of for example FK506 (HAGH).

Currently, only a single example of an AT responsible for the incorporation of a five carbon-CoA unit has been disclosed. In this case the AT displays a different motif at this point, CPTH, in which only the histidine is conserved. The incorporation of a proline residue in the motif may be indicative of an AT specifying a larger substrate. Proline is also found in the motif in ATs that incorporate the larger unusual starter acids as seen in the case of avermectin and soraphen. Residues in and around this area, but lying in the active site of the AT domain define the specificity of the domain towards the substrate chosen.

Motifs that represent hybrids of motifs for malonyl- and methylmalonyl-CoA or methylmalonyl- and ethylmalonyl-CoA were identified. Particularly, epothilone module 3-expected HAFH or YASH (malonyl-CoA or methylmalonyl-CoA specific), found HASH or monensin module 5-expected TAGH (ethylmalonyl-CoA specific), found VAGH. Significantly, in both these cases the products of the PKS are a mixture due to the incorporation of 2 different extender units by

the module containing the hybrid motif, causing formation of monensins A and B and epothilones A and B. However, it is known that substrate supply is a significant determinant of the proportion of monensins A and B formed (Liu, H. and Reynolds, K.A (1999) J. Bact. 181:6806-6813).

Many of the previously-proposed "predictive" motifs are unlikely to be the principal determinant of substrate specificity because they are not located in the active site pocket. A particular requirement of any motif that can serve to distinguish between substrates is that it lies close to the active site and preferably within the substrate binding pocket. In this analysis we consider the substrate binding pocket to be the part of the pocket that binds/recognises the malonyl portion of the acylthioester rather than necessarily the coenzyme A portion. In all probability some of the similarities previously identified by sequence analysis are due to evolutionary conservation rather than a mechanistic requirement. In contrast the residues we have identified lie in or close to the substrate binding pocket. To assess the exact location of the motif in space we compared the protein sequence of ATs derived from Type I PKS with that of *E. coli* fatty acid malonyl-CoA:ACP acyltransferase, for which there is a high resolution X-

ray crystal structure (Serre, L. *et al.*, J. Biol. Chem. (1995) 270:12961-12964). While overall level of sequence similarity between these proteins is low, key residues (and particularly those with mechanistic importance) are conserved and the overall spatial arrangement of amino acids is expected to be conserved. Many groups have used this structure as a model AT and it is well known in the art that conservation of structure can be greater than the level of sequence conservation. Structural analysis showed that the identified motif would lie within the active site pocket opposite the active site serine and the arginine thought to be involved in binding the substrate carboxylate and close enough to the acyltransferase site to interact with the bound substrate side chain. The invariant histidine found in the motif is thought be part of a catalytic triad with the active site serine as is typically found in serine hydrolases (*Serre et al, Supra*). Figure 3 shows the position of the motif loop and important active site residues in the model AT structure.

Broadly the invention concerns modifying an AT domain by changing the four-residue sequence or motif responsible for selecting a substrate so that its specificity is altered. We may also change a small number of other residues close to the active site.

Generally the total number of residues changed is less than 5% of the residues of the AT.

The motif is the four-residue sequence corresponding to the YASH motif found at about residues 334-337 of the AT domain of the first module of DEBS, numbering as shown in Fig. 2. It lies in the active site pocket. It typically starts 80-110, more particularly 90-100, amino acids downstream of the GHS active site motif.

In a preferred embodiment of this invention polyketides of desired structure are produced by the replacement of an existing AT motif on a PKS with an alternative one responsible for selection of an alternative extender or starter unit, or responsible for an altered degree of selectivity (in most cases, increased selectivity). This may be carried out in one or more of the modules encoding a PKS cluster. One type of embodiment is a PKS including two adjoining domains, which are "naturally" adjoining or otherwise coupled domains, wherein the first of them is an AT domain where the four-residue motif has been altered to change its specificity, the AT domain acting to transfer a substrate to the second domain.

In one class of embodiments, this invention provides a PKS multienzyme or part thereof, or nucleic acid (generally DNA) encoding it, said multienzyme or part

comprising a loading module and a plurality of extension modules for the generation of a polyketide, preferably selected from, macrolides, polyethers, or polyenes, wherein the loading or extension modules or at least one thereof contain a modified AT domain adapted to load and transfer an optionally substituted malonyl-CoA residue to (preferably) the ACP. The AT domain is preferably modified to alter its substrate specificity. This AT domain may differ from one naturally found in this position in the module only by the modification of a few amino acids lying in the active site. This modification comprises the exchange of all or part of a motif particularly but not limited to HAFH with YASH or TAGH or vice versa. Optionally, alterations to amino acids outside this sequence, but preferably lying close to the AT active site, are made.

A second class of embodiments provides a method of synthesising polyketides having a desired extension unit at any point around the polyketide molecule by providing a PKS multienzyme incorporating one or more modified AT domains and particularly but not limited to an AT domain possessing the motif HAFH or YASH or TAGH where these motifs replace the existing natural motif. Optionally, alterations to amino acids outside this sequence, but preferably lying close to the AT active site, are made.

A third class of embodiments provides a method of synthesising polyketides having a desired starter unit by providing a PKS multienzyme incorporating a modified AT domain in the loading module and particularly (but not limited to) an AT domain possessing the motif HAFH or YASH or TAGH or a motif incorporating a proline residue where these motifs replace the existing natural motif. Optionally, alterations to amino acids outside this sequence, but preferably lying close to the AT active site, are made. Preferentially, this AT will follow a KSQ domain but other loading systems are known in the art (e.g. AT-ACP). Patent application WO 00/00500 describes some of the known loading systems. The modification of the loading module can be combined with similar modifications in other extension units.

A further class of embodiments provides a method of synthesising polyketides free of natural co-produced analogues and having a desired extender or loading unit by replacing an existing hybrid or alternative protein motif with the sequences HAFH, YASH or TAGH. It is particularly useful to make this alteration in the epothilone or monensin PKS gene cluster.

In still further aspects this invention provides a method of synthesising a mixed population of polyketides by providing a PKS multienzyme incorporating an AT with a

altered or hybrid motif, particularly, but not limited to HASH or VAGH. One particular utility of this method, though not limited to this utility, is the production of combinatorial libraries of compounds.

In a further aspect the PKS containing a modified AT domain may be spliced to a hybrid PKS produced for example as in WO 98/01546 and WO 98/01571 or WO 00/01827 or WO 00/00500. It is particularly useful to link such a modified PKS to gene assemblies that produce novel derivatives of natural polyketides, for example 14-membered macrolides.

Each of these aspects and classes of embodiment may involve providing nucleic acid encoding the polyketide synthase multienzyme and introducing it into a organism where it can be expressed. Suitable plasmids and host cells are described below. The polyketide synthase so produced or portions thereof may be isolated from the host cells by routine methods, though it is usually preferable not to do so. The host cells may also be capable of producing the required acylthioester, eg. by producing ethylmalonyl CoA for example. It may be advantageous to remove the PKS from a strain with a particularly strong supply of an undesired acylthioester or express the altered PKS in a strain specifically chosen to have a strong supply of a particular

acylthioester, or alternatively to develop media or growth conditions to enhance expression of the desired product. Conversely, such techniques could be used to promote formation of mixtures of products if so desired.

It may also be beneficial to supply chemical precursors to the desired acylthioesters in the media e.g. supply diethylethylmalonate or cyclobutane carboxylic acid etc.

The host cells may also be capable of modifying the initial PKS products, e.g. by carrying out all or some of the biosynthetic modifications normal in the production of erythromycin (as shown in figure 4) and for other polyketides. Use may be made of mutant organisms such that some or all of the normal pathways are blocked, e.g. to produce products without one or more "natural" hydroxy groups or methyl groups or sugar groups.

The invention should not be limited to the exact motifs described. We have described some of the known variations within the motif, particularly at residue 1 as can be determined by inspection of Figure 2 or by inspection of similar sequence data. However other modifications can be envisaged; substitution of, for example, the phenylalanine in the malonyl-CoA motif by the similar sized tyrosine may still display the same selectivity. Similarly substitution of the small residue glycine found in the motif responsible for ethylmalonyl-

CoA/other extender incorporation by for example but not limited to alanine. It is well known to those skilled in the art that these and other similar conservative substitutions frequently maintain the same selectivity. Similarly the serine residue found in the motif for incorporation of methylmalonyl-CoA could be substituted by a residue intermediate in size and/or displaying a similar charge distribution.

The invention should not be limited to changes only in this motif. Alterations to other residues around the AT domain may also be required to increase the level of specificity or catalytic efficiency, i.e. to increase the proportion or amounts of the desired products. These residues are preferentially close to the substrate binding pocket. The requirement for these additional alterations will depend on the particular context or change desired. Particular residues to alter can be readily identified by inspection of Figure 2 or other similar sequence analysis data or alternatively by analysis of the structural model.

Residues that may be altered in addition to the motif can be divided into two classes. Some of these residues may have been previously identified in the motifs used to predict the specificity of a motif (ie. Haydock *et al*. (FEBS Lett. (1995) 374:246-248). These

residues are preferentially close to the substrate-binding pocket. These residues should not be limited to the particular examples described.

I) The first class of potential residues to change includes residues close to the motif on the polypeptide chain. A particular example is the residue immediately after the 4 residue motif described in the present invention. In malonyl-CoA specific ATs this residue is generally serine (S), i.e. the protein sequence at this point is generally HAFHS, whereas in methylmalonyl-CoA specific ATs this residue can be S but can also be T, G, or C for example. Thus to change a methylmalonyl-CoA specific AT to a malonyl-CoA specific AT by changing the signature motif it may be beneficial also to ensure that the residue immediately after the motif is an S. Since this residue is close to the motif on the polypeptide chain it lies close to the substrate binding pocket.

II) The second class includes residues that are close to the motif or active site in space. These residues are best identified by reference to the model AT structure described previously or another AT structure that may be subsequently derived. It is known to those skilled in the art that it is possible to thread related protein sequences into an existing structure by using structure molecular modelling or related techniques.

Alternatively, an acylthioester may be modelled into the active site. These are the preferred methods, but often-simple inspection of the existing structure using the highly conserved motifs as a reference point gives a reasonable approximation.

A particular example of a residue close in space to the motif that might be changed is the residue immediately after the GHS active site motif. In methylmalonyl-CoA specific ATs this residue is generally glutamine (Q), i.e. the protein sequence at this point is GHSQ, whereas in malonyl-CoA specific ATs this residue is often V, I or L for example. Thus to change a malonyl-CoA specific AT to a methylmalonyl-CoA specific AT by changing the signature motif it may be beneficial also to ensure that the residue immediately after the GHS motif is a Q. Since this residue is close to the active site serine it lies within the substrate-binding pocket.

A further example of a residue close in space that might be altered is the residue lying three residues downstream of the GQG motif. In methylmalonyl-CoA specific ATs this residue is generally tryptophan (W), i.e. the protein sequence at this point is GQGXXW, whereas in malonyl-CoA specific ATs this residue is often R, H or T for example. Thus to change a malonyl-CoA specific AT to a methylmalonyl-CoA specific AT by

changing the signature motif it may be beneficial also to ensure that this particular residue after the GQG motif is a W. Analysis of the model AT structure shows that the GQG motif lies close to the active site pocket and consequently so does this tryptophan.

A further example of a residue close in space that might be altered is the residue 4 residues downstream from the conserved arginine referred to above, which is believed to stabilise the carboxylate group of the acylthioester substrate. In malonyl-CoA specific ATs this residue downstream of the R is generally methionine (M), i.e. the protein sequence at this point is RXXXMQ. In methylmalonyl-CoA specific ATs this residue is generally I or L, and in ethylmalonyl-CoA specific ATs it is often W. Thus, for example, to change a methylmalonyl-CoA specific AT to a malonyl-CoA specific AT by changing the signature motif it may be beneficial also to ensure that this particular residue is a methionine. Analysis of the model AT structure shows that this residue lies close to the active site pocket.

In further aspects the present invention provides vectors, such as plasmids or phages (preferably plasmids), including nucleic acids as defined in the above aspects and host cells particularly *Saccharopolyspora or Streptomyces* species transformed

with such nucleic acids or constructs. It will be readily apparent to those skilled in the art that there are multiple molecular biological methods for achieving the desired alterations to the AT domain, particularly at the nucleic acid level, e.g. techniques of site directed mutagenesis or directed evolution. Suitable plasmid vectors and genetically engineered cells suitable for expression of PKS genes with modules incorporating an altered AT domain can readily be designed or selected by those skilled in the art. They include those described in WO 98/01546 as being suitable for expression of hybrid PKS genes of Type I. Examples of effective hosts are *Saccharopolyspora erythraea*, *Streptomyces coelicolor*, *Streptomyces avermitilis*, *Streptomyces griseofuscus*, *Streptomyces cinnamonensis*, *Streptomyces fradiae*, *Streptomyces longisporoflavus*, *Streptomyces hygroscopicus*, *Micromonospora griseorubida*, *Streptomyces lasaliensis*, *Streptomyces venezuelae*, *Streptomyces antibioticus*, *Streptomyces lividans*, *Streptomyces rimosus*, *Streptomyces albus*, *Amycolatopsis mediterranei*, and *Streptomyces tsukubaensis*. These include hosts in which SCP2*-derived plasmids are known to replicate autonomously, such as for example *S. coelicolor*, *S.*

avermitilis and *S. griseofuscus*; and other hosts such as *Saccharopolyspora erythraea* in which SCP2*-derived plasmids become integrated into the chromosome through homologous recombination between sequences on the plasmid insert and on the chromosome; and all such vectors which are integratively transformed by suicide plasmid vectors. A plasmid with an int sequence will integrate into a specific attachment site on the host's chromosome.

It is apparent to those skilled in the art that the overall sequence similarity between nucleic acids encoding comparable AT domains from Type I PKSs is sufficiently high and the domain organisation of different Type I PKSs so consistent between different polyketide-producing organisms, that the processes for obtaining novel hybrid polyketides described will be generally applicable to all natural modular Type I PKSs or their derivatives.

The present invention will now be illustrated, but is not intended to be limited, by means of some examples.

Amino acids have been defined throughout by their standard one letter codes as follows. A-alanine, R-arginine, N-asparagine, D-aspartic acid, C-cysteine, Q-glutamine, E-glutamic acid, G-glycine, H-histidine, I-isoleucine, L-leucine, K-lysine, M-methionine, F-

phenylalanine, P-proline, S-serine, T-threonine, W-tryptophan, Y-tyrosine and V-valine.

Brief Description of Drawings

Figure 1 is a diagram showing the functioning of 6-deoxyerythronolide B synthase (DEBS), a modular PKS producing 6-deoxyerythronolide B, a precursor of erythromycin A.

Figure 2 gives the amino acid sequence comparison of the AT domains of representative Type I PKS gene clusters. The motifs GQG, GHS and LPTY are marked at the base of the figure along with the arginine and the motif defined in the invention as defining specificity. The abbreviations used at the side to define the PKS used are: ave: avermectin, debs: erythromycin, epo: epothilone, sor: soraphen, fkb: FK506, rap: rapamycin, tyl: tylosin, mon: monensin, nid: niddamycin, nys: nystatin, rif: rifamycin. The numbers represent the module number. The letter a at the end of the designation indicates malonyl-CoA specific AT, the letter p indicates methylmalonyl-CoA specific AT, and the letter b indicates ethylmalonyl-CoA specific AT. Further types of AT with unusual or ill-defined AT specificity are indicated with letter x. Due to the numbers of sequences considered, in the pileup each section of 50 amino acids

spreads over two pages. The sequences of the monensin ATs are unpublished. They are set out in PCT/GB00/02072.

Figure 3 shows a three-dimensional representation of the active site of the *E. coli* acyltransferase. The spatial arrangement of the motifs described in the text are shown by arrows and the atoms shown in bold.

Figure 4 shows the enzymatic steps that convert 6-deoxyerythronolide B into erythromycin A in *Saccharopolyspora erythraea*.

Figure 5 shows the DNA sequence from the monensin PKS encoding the loading AT used in Example 8.

Modes for Carrying Out the Invention

Example 1

Construction of plasmid pHP41

Plasmid pHP41 is a pCJR24-based plasmid containing the DEBS1 PKS gene comprising a loading module, the first and second extension modules of DEBS and the chain terminating thioesterase. The motif YASH of the AT domain of first module has been altered to HAFH. Plasmid pHP41 was constructed by several intermediate plasmids as follows. Plasmid pD1AT2 (Oliynyk, M. et *al*. Chem. Biol. (1996) 3:833-839) was digested with *Nde*I and *Xba*I. A ~11kbp fragment was isolated by gel electrophoresis and

the DNA purified from the gel. This fragment was ligated into pCJR24 (Rowe, C.J. *et al*. Gene (1998) 216:215-223) that had been linearised by digestion with *Nde*I and *XbaI* and treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E. coli* DH10B cells and individual clones checked for the desired plasmid pCJR26. Plasmid pCJR26 was identified by restriction pattern. pCJR26 was transformed into *E. coli* strain ET12567 (McNeil, D.J. et al. Gene (1992) 111:61-68) and an individual colony grown overnight to isolate demethylated DNA. This DNA was linearised using *Msc*I and *Avr*II and the ~13kb fragment (Fragment A) isolated by gel electrophoresis and purification from the gel.

A DNA segment of the eryAI gene (start nucleotide 45368, end nucleotide 34734) from *S.erythraea* extending from nucleotide 42104 to nucleotide 41542 was amplified by PCR using the following oligonucleotide primers; 5'-TTTTTTTGGCCAGGGTTGGCAGTGGGCGGGCA-3' and 5'-TTTTTACGGCCAGCCGCTTGGCGCGGAT-3'. The DNA from a plasmid designated pCJR65 derived from pCJR24 and DEBS1TE was used as a template. The design of the primers introduced a *Msc*I site at nucleotide 42105 and the second primed across a *Bst*XI site at position 41546. The 574bp PCR

product was treated with T4 polynucleotide kinase and ligated to plasmid pUC18 that had been linearised by digestion with *Sma*I and then treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E. coli* DH10B and individual clones checked for the presence of the desired plasmid pHP39. Plasmid pHP39 was identified by restriction pattern and sequence analysis. Demethylated DNA was produced by transforming *E. coli* strain ET12567 with plasmid DNA. The resulting DNA was linearised by digestion with *Msc*I and *Bst*XI and the resulting 552bp fragment (Fragment B) isolated by gel electrophoresis and purified from the gel. A DNA segment of the eryAI gene from *S.erythraea* extending from nucleotide 41557 to nucleotide 41120 was amplified by PCR using the following oligonucleotide primers; 5'-CGGTGCCTAGGTGCACCGACTCCCAGTCC-3 5'-TTTTTCCAAGCGGCTGGCCGTGGACCACGCGTTCCACTCCTCGCACGTCGAGACGAT-3'. DNA from plasmid pCJR65 was used as a template. The design of the primers introduced an *Avr*II site at nucleotide 41125 and the second primed across a *Bst*XI site at nucleotide 41557 and mutated the amino acid sequence YASH to HAFH (encoded by nucleotides 41537-41526). The 442bp PCR product was treated with T4 polynucleotide kinase and ligated to plasmid pUC18 that

had been linearised by digestion with *Sma*I and then treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E. coli* DH10B and individual clones checked for the presence of the desired plasmid pHP40. Plasmid pHP40 was identified by restriction pattern and sequence analysis. Plasmid pHP40 was linearised by digestion with restriction enzymes *Avr*II and *Bst*XI, and a 427bp fragment (Fragment C) isolated by gel electrophoresis and purified from the gel. Fragments A, B, and C were ligated together and the resulting ligation mixture used to transform electrocompetent *E. coli* DH10B. Individual clones were checked for the presence of an insert derived from DEBS1. The resulting plasmid was designated pHP41. Sequence analysis was used to confirm the clone contained the correct motif HAFH.

Example 2

Construction of *S. erythraea* NRRL2338 JC2/pHP41 and production of triketides

S. erythraea NRRL2338 JC2 contains a deletion of the eryAI, eryAII and eryAIII apart from the TE (Rowe, C.J. *et al*. Gene 216, 215-223). Plasmid pHP41 was used to transform *S. erythraea* NRRL2338 JC2 protoplasts using the

TE as a homology region. Thiostrepton resistant colonies were selected on R2T20 agar containing 40 µg/ml thiostrepton. *S. erythraea* NRRL2338 JC2 (pHP41) was plated onto SM3 agar (see patent application WO 00/01827) containing 40 µg/ml thiostrepton and allowed to grow for 11 days at 30°C. Approximately 1cm^2 of the agar was homogenised and extracted with a mixture of 1.2ml ethyl acetate and 20 µl formic acid. The solvent was decanted and removed by evaporation and the residue dissolved in methanol and analysed by GC/MS. The major products were identified by comparison with authentic standards (Oliynyk, M. *et al*. Chem. Biol. (1996) 3:833-839) as triketide lactones (2S,3R,5R)-2-methyl-3,5-dihydroxy-n-hexanoic δ-lactone (AAP, i.e. Acetate, Acetate, Propionate incorporation) ,(2S,3R,5R)-2-methyl-3,5-dihydroxy-n-heptanoic δ-lactone (PAP), (2R,3S,4S,5R) 2, 4-dimethyl-3,5-dihydroxy-n-heptanoic δ-lactone (PPP)and (2R,3S,4S,5R) 2, 4-dimethyl-3,5-dihydroxy-n-hexanoic δ-lactone (APP). These products were identified as their ammonium adducts corresponding to exact mass 144, 158, 172 and 158. Four products were produced because in this strain, and under the conditions of the experiment the loading module loads both acetate and propionate and the modified AT loads malonyl-CoA and methylmalonyl-CoA.

Only three triketide lactone peaks could be observed in the GC/MS spectra under standard conditions, this was due to the co-elution of the equivalent mass APP and PAP compounds. An isocratic gradient was used to verify this peak was comprised of two components. In further sets of experiments *S. erythraea* JC2 (pHP41) was used to inoculate 5ml TSB containing 5 µg/ml thiostrepton. After three days growth 1.5ml of this culture was used to inoculate 25ml SM3 media containing 5 µg/ml thiostrepton in a 250ml flask. The flask was incubated at 30 °C, 250rpm for 6 days. At this time the supernatant was adjusted to pH3.0 with formic acid and extracted twice with an equal volume of ethyl acetate. The solvent was removed by evaporation and the residue analysed by GC/MS. In each experiment we could identify the 4 products AAP, PAP, PPP and APP but the absolute ratios and quantities were variable, presumably depending on exact media and growth conditions within each flask (figure 6).

Example 3

Construction of *S. erythraea* NRRL2338 (pHP41) and its use to produce 12-desmethyl erythromycin B.

Plasmid pHP41 was used to transform *S. erythraea* NRRL2338 protoplasts. Thiostrepton resistant colonies

were selected on R2T20 agar containing 40 µg/ml thiostrepton. Several clones were tested for the presence of pHP41 integrated into the chromosome by Southern blot hybridisation of their genomic DNA with DIG labelled vector DNA. A clone with a correctly integrated copy of pHP41 was identified in this way. *S. erythraea* NRRL2338 (pHP41) was used to inoculate 5ml TSB containing 5 µg/ml thiostrepton. After three days growth 1.5ml of this culture was used to inoculate 25ml EryP media (see patent application WO 00/00500) containing 5 µg/ml thiostrepton in a 250ml flask. The flask was incubated at 30 ˚C, 250rpm for 6 days. At this time the supernatant was adjusted to pH9.0 with ammonia and extracted twice with an equal volume of ethyl acetate. The solvent was removed by evaporation and the residue analysed by HPLC/MS. A peak of molecular mass m/z (M+H)=704 was observed required for C-12 desmethyl erythromycin B in addition to a peak corresponding to erythromycin A (M+H)=734. Other peaks corresponding to partially processed erythromycin intermediates could be identified.

Example 4

Construction of plasmid pHP048

Plasmid pHP048 is a pCJR24-based plasmid containing the DEBS1 PKS gene comprising a loading module, the first and

second extension modules of DEBS1 and the chain terminating thioesterase. The motif YASH of the AT domain of first module has been altered to HASH. Plasmid pHP048 was constructed by several intermediate plasmids as follows.

A DNA segment of the eryAI gene from *S.erythraea* extending from nucleotide 41557 to nucleotide 41120 was amplified by PCR using the following oligonucleotide primers; 5'-CGGTGCCTAGGTGCACCGACTCCCAGTCC-3' and 5'-TTTTTCCAAGCGGCTGGCCGTGGACCACGCGTCGCACTCCTCGCACGTCGAGACGAT-3'. The DNA from plasmid pCJR65 was used a as template. The design of the primers introduced a *Avr*II site at nucleotide 41125 and the second extended to a *Bst*XI site at nucleotide 41557, also mutated the amino acid sequence YASH (encoded by nucleotides 41537-41526) to HASH. The 442bp PCR product was treated with T4 polynucleotide kinase and ligated to plasmid pUC18 that had been linearised by digestion with *Sma*I and then treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E. coli* DH10B and individual clones checked for the presence of the desired plasmid pHP022. Plasmid pHP022 was identified by restriction pattern and sequence analysis. Plasmid pHP022 was linearised by digestion with restriction enzymes *Avr*II

and *Bst*XI, and the fragment (Fragment D) isolated by gel electrophoresis and purified from the gel. Fragment D was ligated with Fragments A and B described previously and the resulting ligation mixture used to transform electrocompetent *E. coli* DH10B. Individual clones were checked for the presence of an insert derived from DEBS1. The resulting plasmid was designated pHP048. Sequence analysis was used to confirm the clone contained the correct motif HASH.

Example 5

Construction of *S. erythraea* NRRL2338 JC2 (pHP048) and its use to produce triketides

S. erythraea NRRL2338 JC2 contains a deletion of the eryAI, eryAII and eryAIII apart from the TE (Rowe, C.J. *et al*. Gene 216, 215-223). Plasmid pHP048 was used to transform *S. erythraea* NRRL2338 JC2 protoplasts using the TE as a homology region. Thiostrepton resistant colonies were selected on R2T20 agar containing 40µg/ml thiostrepton. *S. erythraea JC2* (pHP048) was used to inoculate 5ml TSB containing 5 µg/ml thiostrepton. After three days growth 1.5ml of this culture was used to inoculate 25ml SM3 media containing 5 µg/ml thiostrepton in a 250ml flask. The flask was incubated at 30 °C,

250rpm for 6 days. At this time the supernatant was adjusted to pH3.0 with formic acid and extracted twice with an equal volume of ethyl acetate. The solvent was removed by evaporation and the residue analysed by GC/MS. A mixture of products were identified as their ammonium adducts corresponding to the AAP, PAP, APP and PPP triketide lactones as described in example 2. In this example, under the media/growth conditions described the PKS with the HASH change is more catalytically active than the HAFH change (example 2) as judged by total amounts of triketide lactone produced, however in this case the modified PKS appears to display lower selectivity towards acetate as judged by the ratio of AAP to PPP triketide lactone.

Example 6

Construction of plasmid pHP47

Plasmid pHP47 is a pCJR24-based plasmid containing the DEBS1 PKS gene comprising a loading module, the first and second extension modules of DEBS1 and the chain terminating thioesterase. The motif YASH of the AT domain of first module has been altered to VAGH. Plasmid pHP47 was constructed by several intermediate plasmids as follows.

A DNA segment of the eryAI gene from *S.erythraea* extending from nucleotide 41557 to nucleotide 41120 was amplified by PCR using the following oligonucleotide primers; 5'-CGGTGCCTAGGTGCACCGACTCCCAGTCC-3' and 5'-TTTTTCCAAGCGGCTGGCCGTGGACGTCGCGGGGCACTCCTCGCACGTCGAGACGAT-3'. The DNA from plasmid pCJR65 was used as a template. The design of the primers introduced a *Avr*II site at nucleotide 41125 and the second extended to a *Bst*XI site at nucleotide 41557, also mutated the amino acid sequence YASH (encoded by nucleotides 41537-41526) to VAGH. The 442bp PCR product was treated with T4 polynucleotide kinase and ligated to plasmid pUC18 that had been linearised by digestion with *Sma*I and then treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E. coli* DH10B and individual clones checked for the presence of the desired plasmid pHP46. Plasmid pHP46 was identified by restriction pattern and sequence analysis. Plasmid pHP46 was linearised by digestion with restriction enzymes *Avr*II and *Bst*XI, and the fragment (Fragment E) isolated by gel electrophoresis and purified from the gel. Fragment E was ligated with Fragments A and B described previously and the resulting ligation mixture used to transform electrocompetent *E. coli* DH10B. Individual clones were

checked for the presence of an insert derived from DEBS1. The resulting plasmid was designated pHP47. Sequence analysis was used to confirm the clone contained the correct motif VAGH.

Example 7

Construction of plasmid pLS007

Plasmid pLS007 contains the crotonyl-CoA reductase (CCR) gene from *S. cinnamonensis* that is believed to influence the level of ethylmalonyl-CoA within the cell. Plasmid pSG142 (Gaisser *et al*. Mol. Microbiol. (2000) 36 391-401) places genes under the control of the actI promoter and can be used to integrate either in the right hand side of the erythromycin gene cluster or in the act promoter region of a previously transformed actinomycete. Two oligonucleotide primers; 5'-GGCAAACATATGAAGGAAATCCTGGACGCG-3' and 5'-TCCGCGGATCCTCAGTGCGTTCAGATCAGTGC-3' were used to amplify the *S. cinnamonensis* CCR gene using genomic DNA as template. The design of the primers incorporated *Nde*I and *Bam*HI restriction sites to facilitate cloning. The 1.4kb PCR product was isolated by gel electrophoresis and purified from the gel and ligated with pSG142 that had been digested with NdeI and BglII. The resulting

ligation mixture was used to transform electrocompetent *E. coli* DH10B cells. Plasmid pLS003 was identified by restriction analysis and sequencing to ensure errors were not introduced during amplification. A discrepancy with the published sequence was identified. However, further analysis by comparison with other published CCR protein sequences indicated pLS003 was correct. Plasmid pLS003 was digested with *Nde*I and *Xba*I and the resulting 4.5kb fragment (fragment F) isolated by gel electrophoresis and purified from the gel. This fragment was ligated to pLSB2 a derivative of pKC1132 containing the actI/actII promoter region behind an *Nde*I site. Plasmid pLSB2 was digested with *Nde*I and *Xba*I and the resulting ~4kb fragment (Fragment G) purified by gel electrophoresis and purified from the gel. Fragments F and G were ligated together and the resulting ligation mixture was used to transform electrocompetent *E. coli* DH10B cells. Plasmid pLS007 was identified by restriction analysis.

Example 8

Construction of *S. erythraea* NRRL2338 JC2 (pHP47/pLS007) and its use to produce triketides

S. erythraea NRRL2338 JC2 contains a deletion of the eryAI, eryAII and eryAIII apart from the TE (Rowe, C.J.

et al. Gene 216, 215-223). Plasmid pHP47 was used to transform *S. erythraea* NRRL2338 JC2 protoplasts using the TE as a homology region. Thiostrepton resistant colonies were selected on R2T20 agar containing 40 µg/ml thiostrepton. PLS007 was used to transform protoplasts of *S. erythraea* NRRL2338 JC2 (pHP47), thiostrepton and apramycin resistant clones were selected on R2T20 agar containing 40 µg/ml thiostrepton and 50 µg/ml apramycin plus 10mM magnesium chloride and the resistance markers verified by plating on tapwater media containing the same antibiotics. *S. erythraea* NRRL2338 JC2 (pHP47/pLS007) was used to inoculate 5ml TSB containing 5 µg/ml thiostrepton and 50 µg/ml apramycin. After three days growth 1.5ml of this culture was used to inoculate 25ml SM3 media containing 5 µg/ml thiostrepton and 50 µg/ml apramycin in a 250ml flask. The flask was incubated at 30°C, 250rpm for 6 days. At this time the supernatant was adjusted to pH3.0 with formic acid and extracted twice with an equal volume of ethyl acetate. The solvent was removed by evaporation and the residue analysed by GC/MS. In this experiment amounts of triketide product were lower but a mixture of products could be identified as their ammonium adducts corresponding to exact masses 158 172 and 186.

Example 9

Construction of *S. erythraea* NRRL2338 (pHP47) and its use to produce erythromycins.

Plasmid pHP47 was used to transform *S. erythraea* NRRL2338 protoplasts. Thiostrepton resistant colonies were selected on R2T20 agar containing 40 µg/ml thiostrepton. *S. erythraea* NRRL2338 (pHP47) was used to inoculate 5ml TSB containing 5 µg/ml thiostrepton. After three days growth 1.5ml of this culture was used to inoculate 25ml EryP media containing 5 µg/ml thiostrepton in a 250ml flask. The flask was incubated at 30°C, 250rpm for 6 days. At this time the supernatant was adjusted to pH9.0 with ammonia and extracted twice with an equal volume of ethyl acetate. The solvent was removed by evaporation and the residue analysed by HPLC/MS. Peaks of mass m/z (M+H)=734 corresponding to erythromycin A were observed.

Example 10

Construction of plasmid pSGK051

Plasmid pSGK051 is a pPFL43 based plasmid (WO 00/00500). The motif HAFH of the AT domain of the loading domain has been altered to YASH. Plasmid pSGK051 was constructed by several intermediate plasmids as

follows.

Plasmid pPFL43 was linearised by digestion with restriction enzymes *Nco*I and *Not*I and a 858bp fragment (Fragment Q) isolated by gel electrophoresis and purified from the gel.

A DNA segment of the monensin loading domain from nucleotide 16360-17366 (see figure 5 and PCT/GB00/02072) was amplified by PCR using the following oligonucleotide primers; 5'-GGGGACGCGGCCGCAAGGCCCACCACCTGAAGGTCAGCTACGCCTCCCACTCCCCGCACATGGACCCCAT-3' and 5'-GGCTAGCGGGTCCTCGTCCGTGCCGAGGTCA-3'. The design of the primers amplified across a *Not*I site at nucleotide 16367 and changed the amino acid sequence HAFH to YASH at nucleotides 16398-16409, the second introduced a *Nhe*I site equivalent to that in pPFL43. The DNA from plasmid pPFL43 was used as a template. The 1006bp PCR product was treated with T4 polynucleotide kinase and ligated to plasmid pUC18 that had been linearised by digestion with *Sma*I and treated with alkaline phosphatase. The ligation mixture was used to transform electrocompetent *E. coli* DH10B and individual clones checked for the presence of the desired plasmid pCSAT9. Plasmid pCSAT9 was identified by restriction pattern and sequence analysis. Plasmid

pCSAT9 was linearised by digestion with restriction enzymes *Not*I and *Nhe*I and a 995bp fragment (Fragment R) isolated by gel electrophoresis and purified from the gel. Plasmid pPFL43 was digested with *Nco*I and *Nhe*I to remove a 1.8kb fragment and the larger fragment (Fragment S) isolated by gel electrophoresis and purified from the gel. Fragments Q, R and S were ligated together and the resulting ligation mixture used to transform electrocompetent *E. coli* DH10B. Individual clones were checked for the desired plasmid pSGK051. The resulting plasmid was analysed by restriction digest and sequenced to confirm the presence of the correct motif YASH.

Example 11

Construction of *S. erythraea* NRRL2338 JC2/pSGK051 and production of triketides

Plasmid pSGK051 was used to transform *S. erythraea* NRRL2338 JC2 protoplasts using the TE as a homology region. Thiostrepton resistant colonies were selected on R2T20 agar containing 40 µg/ml thiostrepton. *S. erythraea* NRRL2338 JC2 (pSGK051) was plated onto R2T20 agar containing 40 µg/ml thiostrepton and allowed to grow for 11 days at 30°C. Approximately 1cm^2 of the agar was homogenised and extracted with a mixture of 1.2ml ethyl

acetate and 20 µl formic acid. The solvent was decanted and removed by evaporation and the residue dissolved in methanol and analysed by GC/MS. The major products were identified by comparison with authentic standards as triketide lactones (2S,3R,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-heptanoic δ-lactone and (2S,3R,4S,5R)-2,4-dimethyl-3,5-dihydroxy-n-hexanoic δ-lactone.

Example 12

Construction of *S. erythraea* NRRL2338 (pSGK051) and its use to produce erythromycins.

Plasmid pSGK051 was used to transform *S. erythraea* NRRL2338 protoplasts. Thiostrepton resistant colonies were selected on R2T20 agar containing 40 µg/ml thiostrepton. *S. erythraea* NRRL2338 (pSGK051) was plated onto R2T20 agar containing 40 µg/ml thiostrepton and allowed to grow for 10 days at 30°C. Approximately 2cm^2 of the agar was homogenised and extracted with a mixture of 1.2ml ethyl acetate and 20 µl dilute ammonia. The solvent decanted and was removed by evaporation and the residue analysed by HPLC/MS. Peaks of mass m/z (M+H)=734 and 720 could be observed alongside likely products of incomplete processing. Comparison to authentic standards proved the compounds produced were erythromycin A and 13-

methyl erythromycin A.

CLAIMS:

1. A method of synthesising a compound whereof at least a portion is the product of a polyketide synthase (PKS) enzyme complex or is derived from such a product, said PKS enzyme complex including at least one acyltransferase (AT) domain; said method comprising the steps of (i) providing said PKS enzyme complex in which said AT domain has been altered to change selectively a minor proportion of amino acid residues, the altered residue(s) comprising one or more residues of one or more motifs which are present in the active site pocket of the AT domain and which influence the substrate specificity of the AT domain, the alteration affecting the substrate specificity; and (ii) effecting synthesis by means of said PKS enzyme complex to produce a compound or mixture of compounds different from what could have been produced by means of a PKS enzyme in which said AT domain had not been altered.

2. A method according to claim 1 wherein said motif comprises a four-residue sequence corresponding to the YASH motif of the AT domain of the first module of DEBS.

3. A method of synthesising a compound whereof at least a portion is the product of a polyketide synthase

(PKS) enzyme complex or is derived from such a product, said PKS enzyme complex including at least one acyltransferase (AT) domain; said method comprising the steps of (i) providing said PKS enzyme complex in which said AT domain has been altered to change selectively a minor proportion of amino acid residues, the altered residue(s) comprising one or more residues of a motif which influences the substrate specificity of the AT domain and which comprises a four-residue sequence corresponding to the YASH motif of the AT domain of the first module of DEBS, the alteration affecting the substrate specificity; and (ii) effecting synthesis by means of said PKS enzyme complex to produce a compound or mixture of compounds different from what could have been produced by means of a PKS enzyme in which said AT domain had not been altered.

4. A method according to claims 1, 2 or 3 wherein said motif was located by a) determining the sequence of the AT domain and b) performing sequence alignment with a plurality of sequences of other AT domains.

5. A method according to any preceding claim wherein the PKS enzyme complex is at least part of a modular type I PKS enzyme complex.

6. A method according to any preceding claim wherein said alteration of the AT domain affects less than 5% of the residues.

7. A method according to any preceding claim wherein said alteration alters a motif selected from XAFH, XASH, and XAGH and/or creates such a motif.

8. A method according to claim 7 wherein the motif is XAGH and X is selected from F, T, V and H.

9. A method according to claim 7 wherein the motif is XAFH and X is H.

10. A method according to claim 7 wherein the motif is XASH and X is selected from Y,H,W and V.

11. A method according to any of claims 1-10 wherein said alteration produces or alters a motif containing proline.

12. A method according to any preceding claim wherein in addition to the alteration to one or more residues of said motif(s), one or more additional residues in or adjacent the substrate binding pocket have been altered.

13. A method according to claim 12 wherein said additional altered residue(s) comprise one or more of a) the residue immediately downstream of the motif, b) the residue three residues downstream from the GQG motif, c) the residue immediately downstream of the GHS motif, and

d) the residue four residues downstream of the conserved arginine residue.

14. A method according to any preceding claim wherein the alteration produces a motif specific for malonyl-CoA and the motif is followed by S which was produced by alteration if not already present.

15. A method according to any of claims 1-13 wherein the alteration produces a motif specific for methylmalonyl-CoA and the motif is followed by S, G, C or T which was produced by alteration if not already present.

16. A method according to any of claims 1-13 wherein the alteration produces a motif specific for methylmalonyl-CoA, and the residue following the GHS motif in the active site is Q which was produced by alteration if not already present.

17. A method according to any of claims 1-13 wherein the alteration produces a motif specific for malonyl-CoA, and the residue following the GHS motif in the active site is V, I or L which was produced by alteration if not already present.

18. A method according to any of claims 1-13 wherein the alteration produces a motif specific for methylmalonyl-CoA, and the residue 3 residues downstream

of the GQG motif is W which was produced by alteration if not already present.

19. A method according to any of claims 1-13 wherein the alteration produces a motif specific for malonyl-CoA, and the residue 3 residues downstream of the GQG motif is R, H or T which was produced by alteration if not already present.

20. A method according to any of claims 1-13 wherein the alteration produces a motif specific for malonyl-CoA and the residue 4 residues downstream of the conserved R as found as residue 252 in the first module of DEBS is M which was produced by alteration if not already present.

21. A method according to any of claims 1-13 wherein the alteration produces a motif specific for methylmalonyl-CoA and the residue 4 residues downstream of the conserved R as found as residue 252 in the first module of DEBS is I or L which was produced by alteration if not already present.

22. A method according to any of claims 1-13 wherein the alteration produces a motif specific for ethylmalonyl-CoA and the residue 4 residues downstream of the conserved R as found as residue 252 in the first module of DEBS is W which was produced by alteration if not already present.

23. A method according to any preceding claim wherein the AT domain has an active site with a GHS motif, and said motif which is altered starts 80-110 residues downstream of said GHS motif.

24. A method according to any preceding claim wherein said step (i) of providing said PKS enzyme complex comprises providing a nucleic acid sequence encoding said complex and effecting expression thereof.

25. A method according to claim 24 wherein expression is effected is an organism capable of producing polyketides.

26. A method according to claim 24 or claim 25 wherein said nucleic acid sequence has been subjected to site directed mutagenesis so that it encodes said altered AT domain.

27. A method according to claim 24, 25 or 26 wherein the AT domain prior to alteration is naturally expressed in a first organism and the altered AT is expressed in a second organism which is better able than the first organism to supply a substrate for which the alteration has increased specificity and/or which is less well able than the first organism to supply a substrate for which the alteration has reduced specificity.

28. A method according to any preceding claim wherein said PKS includes said AT domain and a second

domain which is naturally coupled thereto prior to the alteration thereof to receive a substrate transferred to it by the AT; and the alteration causes the AT to act to transfer a different substrate to the second domain.

29. A method according to any preceding claim wherein said PKS includes said AT domain and its natural cognate ACP domain which, prior to the alteration, is adapted to receive a substrate transferred to it by the AT; and the alteration causes the AT to act to transfer a different substrate to said cognate ACP domain.

30. A method according to any preceding claim wherein said PKS including the altered AT domain is spliced to a hybrid PKS.

31. A polyketide compound or derivative thereof or compound whereof a portion is a polyketide or derivative thereof, which compound is obtainable by a method according to any preceding claim wherein the compound differs from a compound resulting from synthesis effected by means of said PKS enzyme complex without the alteration of said AT domain.

32. Nucleic acid encoding a PKS enzyme complex including an altered AT domain as defined in any of claims 1-30.

33. A vector including a nucleic acid according to claim 32.

34. A host organism containing nucleic acid according to claim 32 and able to express the PKS enzyme complex.

35. A host organism according to claim 34 which is adapted to synthesise a compound whereof at least a portion is a polyketide resulting from the action of the PKS enzyme complex.

36. A method of synthesising a polyketide synthase (PKS) enzyme complex, said PKS enzyme complex including at least one acyltransferase (AT) domain; said method comprising altering said AT domain to change selectively a minor proportion of amino acid residues, the altered residue(s) comprising one or more residues of one or more motifs which are present in the active site pocket of the AT domain and which influence the substrate specificity of the AT domain, the alteration affecting the substrate specificity.

37. A method according to claim 36 wherein said motif comprises a four-residue sequence corresponding to the YASH motif of the AT domain of the first module of DEBS.

38. A method of synthesising a polyketide synthase (PKS) enzyme complex, said PKS enzyme complex including at least one acyltransferase (AT) domain; said method comprising altering said AT domain to change selectively

a minor proportion of amino acid residues, the altered residue(s) comprising one or more residues of a motif which influences the substrate specificity of the AT domain and which comprises a four-residue sequence corresponding to the YASH motif of the AT domain of the first module of DEBS, the alteration affecting the substrate specificity.

39. A PKS enzyme complex as produced by the method of claims 36, 37 or 38.

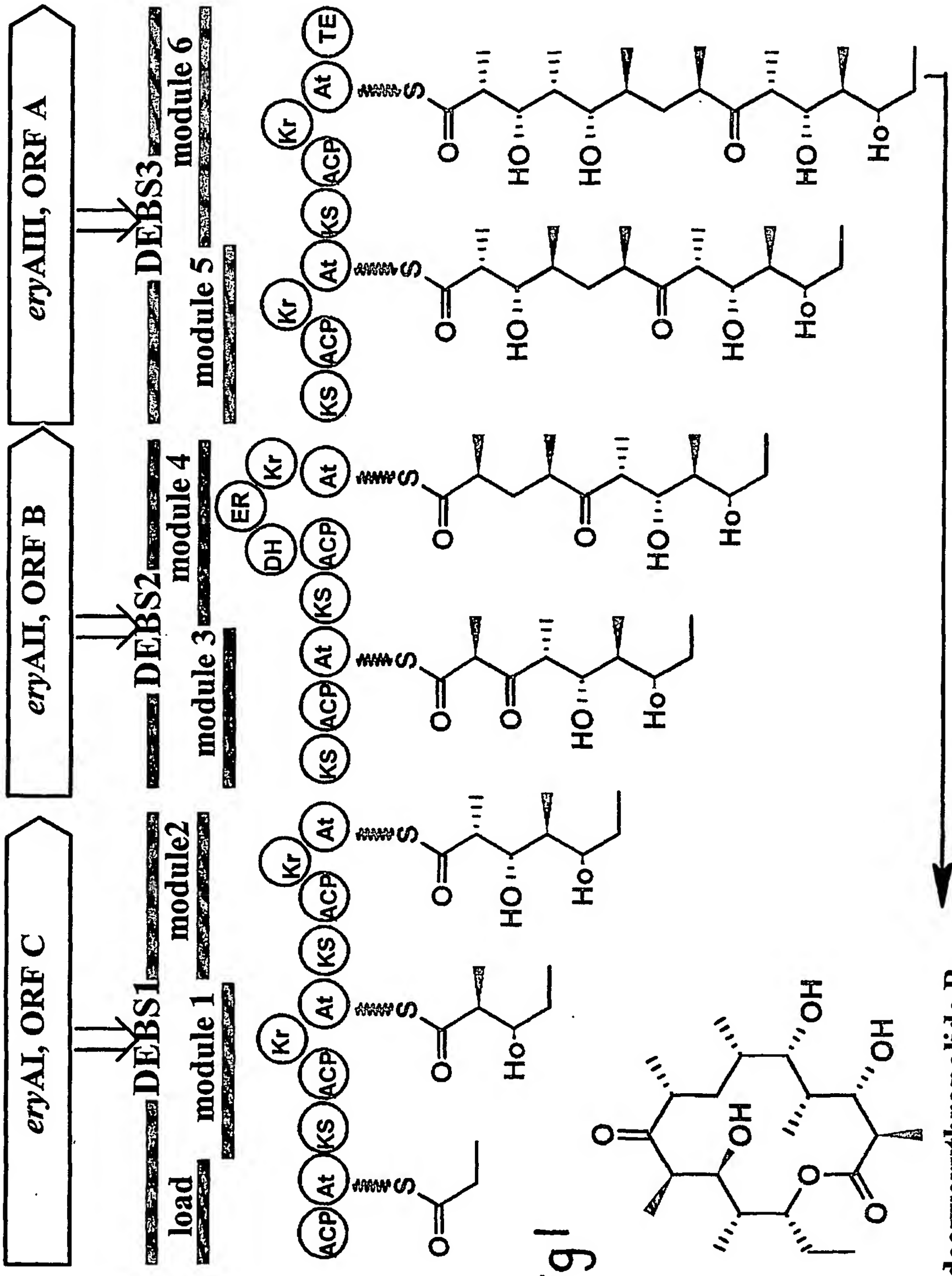


Fig1

6-deoxyerythronolide B

```
                1                                                   50
     atave00x   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   atdebs00p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atepo06p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atepo07p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atepo01p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atepo05p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atsora1x   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atfkb01p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atfkb09p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atrap03p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atrap06p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atrap04p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~PLVI
     atrap13p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~A EEAQPVETPV VASDVLPLVI
     atrap01p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atrap07p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~PV VASELVPLVI
     atrap10p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atfkb04x   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atty104p   ~~~~~~~~~~ ~~~~~~~~VV REAAGRLAEV VEAGGVGLAD VAVTMAGRSR
     atty106p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~GRLAEV VEAGGVGLAD VAVTMAGRSR
     atty101p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~MAGRSR
     atty102p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atty100p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~D VAVTMADRSR
     atnid05b   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atty105b   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~AAL REQSTRLLND
     atnid06x   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   atdebs01p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atmon02p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atmon10p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atmon04p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atmon07p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atmon11p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atmon12p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atmon05b   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atmon01p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   atdebs02p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   atdebs06p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atave01p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atave07p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atave06p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atave09p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atnys01p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atnys11p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atrif05p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atrif07p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atrif08p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atrif10p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atrif03p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atrif06p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atrif04p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atrif01p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atnys02p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atfkb02p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atave11p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   atdebs03p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     atnid04p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   atdebs05p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
   atdebs04p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
```

Fig 2a

```
atave02a  LFAFQVALHR LLTDGYHITP HYYAGHSLGE ITAAHLAGIL TLTDATTLIT
atave05a  LFAFQVALHR LLTDGYHITP HYYAGHSLGE ITAAHLAGIL TLTDATTLIT
atave04a  LFAFQVALHR LLTDGYHITP HYYAGHSLGE ITAAHLAGIL TLTDATTLIT
atave08a  LFAFQVALHR LLTDGYHITP HYYAGHSLGE ITAAHLAGIL TLTDATTLIT
atave03a  LFAFQVALHR LLTDGYHITP HYYAGHSLGE ITAAHLAGIL TLTDATTLIT
atrap02a  LFALQVALFG LL.ESWGVRP DAVVGHSVGE LAAGYVSGLW SLEDACTLVS
atrap11a  LFALQVALFG LL.ESWGVRP DAVIGHSVGE LAAAYVSGVW SLEDACTLVS
atrap08a  LFALQVALFG LL.ESWGVRP DAVVGHSVGE LAAGYVSGLW SLEDACTLVS
atrap12a  LFAMQVALFG LL.ESWGVRP DAVIGHSVGE LAAAYVSGVW SLEDACTLVS
atrap05a  LFALQVALFG LL.ESWGVRP DAVVGHSVGE LAAGYVSGLW SLEDACTLVS
atrap09a  LFALQVALFG LL.ESWGVRP DAVIGHSVGE LAAAYVSGLW SLEDACTLVS
atfkb03a  VFALQVALSA QL.DAWGVRP DVLVGHSIGE LAAAYVAGVW SLDDATELVS
atfkb07x  HFAHQIALTA LL.RSWGITP HAVIGHSLGE ISAACAAGVL SIGDASALLA
atfkb08x  LFAHQAAFTA LL.RSWDITP HAVIGHSLGE ITAAYAAGIL SLDDACTLIT
atnid01a  LFALQTALYR TL.TARGTQA HLVLGHSVGE ITAAHIAGVL DLPDAARLIT
atnid03a  LFALQTALYR TL.TARGTQA HLVLGHSVGE ITAAHIAGVL DLPDAARLIT
atnid02a  LFALQTALYR TL.TAHGTQA HLVLGHSVGE ITAAHIAGVL DLPDAARLIT
atnid00a  LFALQTALYR TL.TARGTQA HLVLGHSVGE ITAAHIAGVL DLPDAARLIT
atfkb10a  LFTLEVALLR LL.EHWGVRP DVVVGHSVGE VTAAYAAGVL TLADATTLIV
atrap14a  IFAMEAALFG LL.EDWGVRP DFVAGHSIGE ATAAYASGML SLENVTTLIV
atmon06a  LFALQVGLAR LW.ESVGVRP DVVLGHSIGE IAAAHVAGVF DLADACRVVG
atmon08a  LFALQVGLAR LW.ESVGVRP DVVLGHSIGE IAAAHVAGVF DLADACRVVG
atmon09a  LFALQVGLAR LW.ESVGVRP DVVLGHSIGE IAAAHVAGVF DLADACRVVG
atepo02a  LFAVEYALTA LW.RSWGVEP ELLVGHSIGE LVAACVAGVF SLEDGVRLVA
atepo03x  LFTVEYALTA LW.RSWGVEP ELVAGHSAGE LVAACVAGVF SLEDGVRLVA
atepo08a  LFALEYALAA LF.RSWGVEP ELVAGHSLGE LVAACVAGVF SLEDAVRLVV
atepo00a  LFTFEYALAA LW.RSWGVEP ELVAGHSIGE LVAACVAGVF SLEDAVFLVA
atepo04a  LFALEYALAA LW.RSWGVEP HVLLGHSIGE LVAACVAGVF SLEDAVRLVA
atnid07a  LFAVETALFR LF.ESWGLMP DVLLGHSIGG LAAAYAAGVF SSADAVRLVA
atty107a  LFAVEVALHR LL.EHWGMRP DLLLGHSVGE LAAAHVAGVL DLDDACALVA
atsor02a  LFALEVALFQ LL.QSFGLKP ALLLGHSIGE LVAAHVAGVL SLQDACTLVA
atsorb1a  LFALEVALFE LL.QSFGLKP ALLLGHSIGE LVAAHVAGVL SLQDACTLVA
atnys09a  LFAVEVALYR LI.ESFGVRP DHLAGHSVGE IVAAHLAGVL SLADAATLVA
atnys12a  LFAVEVALFR LL.TSWGLTP DYLAGHSVGE LAAAHVAGVL SLDDACTLVA
atnys16a  LFAVEVALFR LV.ASWGVGP EFVAGHSVGE IAAAHVAGVF SLVDACRLVV
atnys17a  LFAVEVALFR LV.ASWGVGP EFVAGHSVGE IAAAYVAGVF SLVDACRLVV
atnys03a  LFAVEVALYR LV.ASLGVTP DFVGGHSIGE LAAAHVAGVL SLEDACTLVA
atnys15a  LFAVEVALYR LI.ESWGVAP DFVAGHSIGE IAAAHVAGVF SLEDACTLVA
atnys07a  LFAIEVALFR LV.ESWGVRP DFVAGHSIGE IAAAHVAGVF SLEDACRLVA
atnys08a  LFAVEVALFR LV.ERWGVRP DFVAGHSIGE IAAAHVAGVF SLEDACRLVA
atnys05a  LFAVEVALFR LV.ESWGVRP DFVAGHSIGE IAAAHVVGVF SLEDACRLVA
atnys06a  LFAIEVALFR LV.ESWGVRP DFVAGHSIGE IAAAHVVGVF SLEDACRLVA
atnys04a  LFAIEVALFR LL.EAWGITP DFVAGHSIGE IAAAHVAGVL SLGDACRLVV
atnys14a  LFAVEVALYR LI.ESWGVRP DFVAGHSVGE LAAAHVAGVL SLDDACRLVA
atnys00a  LFAVEVALHR LV.ASLGVTP DFVGGHSVGE IAAAHVAGVL SLEDACRLVA
atnys10a  LFAVEVALFR LV.ESWGVRP DFVAGHSIGE IAAAHVAGVL TLEDACRLVA
atnys18a  LFAVEVALYR LL.ASWGIRP DHVTGHSIGE ITAAHVAGVL TLADACTLVA
atnys13a  LFAVEVALFR LA.ESWRLTP DFVAGHSIGE IAAAHVAGVF SLEDACTLVA
atave10a  LFAFEVALFR LL.ETWGLTP DYVLGHSVGE LAAAHVAGML CLADAVALVV
atrif02a  LFAVETALFR LF.ESWGVRP GLLAGHSIGE LAAAHVSGVL DLADAGELVA
atmon03a  LFALEVALYR QV.TSFGIAP SHLTGHSVGE IAAAHVAGVF SLADACTLVA
atave12a  LFAVQVALFR HL.ERLGVRA DFVAGHSIGE LAAAHVAGVL PLAAACRLVA
atrif09a  LFAVESALFR LA.ESWGVRP DVVLGHSIGE ITAAYAAGVF SLPDAARIVA
atmon00a  LFAIETSLYR LA.ASFGLKP DYVLGHSVGE IAAAHVAGVL SLPDASALVA
atty103a  LFALQTALFR LA.EHHGLRA EALCGHSVGE IAAAHAAGVL TLPDAARLVA
                                  ***                        GHS
```

Fig2j

```
          251                                                    300
 atave00x LWSQAQTT.L AGTGALVSVA ATPDELLPRI APWTEDN.PA RLAVAAVNGP
atdebs00p LWSREMIP.L VGNGDMAAVA LSADEIEPRI ARWDDD.... .VVLAGVNGP
 atepo06p RRSRLL.RRI SGQGEMALVE LSLEEAEAAL .....RGHEG RLSVAVSNSP
 atepo07p RRSRLL.RRI SGQGEMALVE LSLEEAEAAL .....RGHEG RLSVAVSNSP
 atepo01p RRSRLL.RRI SGQGEMAVTE LSLAEAEAAL .....RGYED RVSVAVSNSP
 atepo05p RRSLLL.RRI SGQGEMAVVE LSLAEAEAAL .....LGYED RLSVAVSNSP
 atsora1x AYGRII.RKL RGKGGMGLVA LSWEDAGKEL .....TGYEG RLFRAIEHSA
 atfkb01p LRSQAIAARL AGRGAMASIA VPASAVE... .......TVE GVWIAARNGP
 atfkb09p LRSQTIAAHL AGRGAMASIA LPATAVE... .......TVE GVWVAARNGP
 atrap03p LRSQAIARGL AGRGAMASVA LPAQDVE... .......LVD GAWIAAHNGP
 atrap06p LRSEAIARGL AGRGAMASVA LPAQDVE... .......LVD GAWIAAHNGP
 atrap04p LRSQAIARGL AGRAAMASVA LPAHEIE... .......LVD GAWIAAHNGP
 atrap13p LRSQAIARGL AGRGAMASVA LPAQDVE... .......LVD GAWIAAHNGP
 atrap01p LRSQVIARGL AGRGAMASVA LPAQDVE... .......LVD GAWVAARNGP
 atrap07p LRSQAIARGL AGRGAMASVA LPAHEIE... .......LVD GAWIAAHNGP
 atrap10p LRSQAIARGL AGRGAMASVA LPAQDVE... .......LVD GAWIAAHNGP
 atfkb04x LRSALLVREL AGRGAMGSIA FAA..AA... .......RID GVWVAGRNGT
 atty104p LRAGLIGRYL AGRGAMAAVP LPAGEVEAGL AKWPG..... .VEVAAVNGP
 atty106p LRAGLIGRYL AGRGAMAAVP LPAGEVEAGL AKWPG..... .VEVAAVNGP
 atty101p LRAGLIGRYL AGRGAMAAVP LPAGEVEAGL AKWPG..... .VEVAAVNGP
 atty102p LRAGLIGRYL AGRGAMAAVP LPAGEVEAGL AKWPG..... .VQVAAVNGP
 atty100p LRAGLIGRYL AGRGAMAAVP LPAGEVEAGL AKWPG..... .VEVAAVNGP
 atnid05b LRSRAWLG.L AGKGGMVAVP MPAEELRPRL .....VTWGD RLAVAAVNSP
 atty105b LRSRAWLT.L AGKGGMAAVS LPEARLRERI .....ERFGQ RLSVAAVNSP
 atnid06x GRSRLWGR.L AGNGGMLAVM APAERIRELL .....EPWRQ RISVAAVNGP
atdebs01p LRSRVIAT.M PGNKGMASIA APAGEVRARI .........GD RVEIAAVNGP
 atmon02p VRSDAL.RQL QGHGDMASLS TGAEQAAELI GDRPG..... .VVVAAVNGP
 atmon10p VRSDAL.RRL QGHGDMASLS TGAEQAAELI GDRPG..... .VVVAAVNGP
 atmon04p VRSDAL.RQL QGHGDMASLG TGAEQAAELI GDRPG..... .VVVAAVNGP
 atmon07p VRSDAL.RQL MGQGDMASLG ASSEQAAELI GDRPG..... .VCIAAVNGP
 atmon11p VRSDAL.RQL QGHGDMASLS TGAEQAAELI GDRPG..... .VVVAAVNGP
 atmon12p VRSDAL.RQL MGQGDMASLG AGSEQVAELI GDRPG..... .VCVAAVNGP
 atmon05b VRSVLL.RQL SGRGGMASLG MGQEQAADLI DGHPG..... .VVVAAVNGP
 atmon01p LRSRAL.RQL SGGGAMASLG VGQEQAAELV EGHPG..... .VGIAAVNGP
atdebs02p RRSRAV.RAV AGRGSMLSVR GGRSDVEKLL ADDS...WTG RLEVAAVNGP
atdebs06p LRAKAL.RAL AGKGGMVSLA APGERARALI A..P...WED RISVAAVNSP
 atave01p LRSRALAA.V RGRGGMASVP LPAQEVEQLI ....GERWAG RLWVAAVNGP
 atave07p LRSRALAA.V RGRGGMASVP LPAQEVEQLI ....GERWAG RLWVAAVNGP
 atave06p LRSRALAA.V RGRGAMASLP LPAQDVQQLI ....SERWEG QLWVAALNGP
 atave09p LRSQALAA.V RGRGAMVSLP LPAQDVQQLI ....SERWEG QLWVAALNGP
 atnys01p LRSQAIGRAL AGRGGMMSVA LSVDVLEPRL VE.....FEG RVSVAAVNGP
 atnys11p LRSQAIGRAL AGRGGMMSVA LSVDVLEPRL VE.....FEG RVSVAAVNGP
 atrif05p LRSQAIAAEL SGRGGMASIQ LSHDEVAARL AP.....WAG RVEIAAVNGP
 atrif07p LRSQAIAARL SGRGGMASVA LSEDEANARL GL.....WDG RIEVAAVNGP
 atrif08p LRSQAIAAKL AGRGGMASVA LSEEDAVARL RH.....WAD RVEVAAVNSP
 atrif10p LRSQAIAAKL SGRGGMASVA LGEADVVSRL ........AD GVEVAAVNGP
 atrif03p LRSQAIAGEL AGRGGMASVA LSEEDAVARL TP.....WAN RVEVAAVNSP
 atrif06p LRSQAIATRL AGRGGMASVA LSEEDATAWL AP.....WAD RVQVAAVNSP
 atrif04p LRSQAIAASL AGRGGMASVA LSEEDATARL EP.....WAG RVEVAAVNGP
 atrif01p LRSQAIAAEL SGRGGMASVA LGEDDVVSRL ........VD GVEVAAVNGP
 atnys02p LRSQALP.QL SGRGGMMSVS APVERVTALL AP.....WQE ALSVAAVNGP
 atfkb02p LRSRLVATER AGHGGMVSVP PADFDAAA.. .......WAG RLEVAAVNGP
 atave11p LRSQALA.AL AGQGAMASVG LPVEKLEPRL A.....TWGD RLVIAAVNGA
atdebs03p GRSRLM.RSL SGEGGMAAVA LGEAAVRERL RPWQ.....D RLSVAAVNGP
 atnid04p LRSQLIAREL AGRGSMASVA LAAADVESRL AGAEAGGGVR DVEIAAVNGP
atdebs05p VRSRVL.RRL GGQGGMASFG LGTEQAAERI .....GRFAG ALSIASVNGP
atdebs04p LRSQVL.REL DDQGGMVSVG ASRDELETVL A.....RWDG RVAVAAVNGP
```

Fig 2k

```
atave02a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atave05a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atave04a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atave08a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atave03a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atrap02a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atrap11a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atrap08a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~PPTQPADNA VIERAPEWLP
atrap12a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atrap05a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atrap09a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~D DVRPADAPVV ASVMASELVP
atfkb03a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atfkb07x  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atfkb08x  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnid01a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnid03a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnid02a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnid00a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atfkb10a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atrap14a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atmon06a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atmon08a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atmon09a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atepo02a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atepo03x  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atepo08a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atepo00a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atepo04a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnid07a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atty107a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~LR DHLSRTPGAR
atsor02a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atsorb1a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys09a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys12a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys16a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys17a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys03a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys15a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys07a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys08a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys05a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys06a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys04a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys14a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys00a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys10a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys18a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys13a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atave10a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atrif02a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atmon03a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atave12a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atrif09a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atmon00a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atty103a  SVPAGEPPAA GRPEDTGGAW TVSGRGPAAL RAQAARLYDA LTGTGTGTGQ
```

Fig 2b

```
            51                                                  100
atave00x    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~VQR MDGGEEPRPA
atdebs00p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~VADGRPH
atepo06p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~AAAQGHTP
atepo07p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~SSREALRGA LSAAAQGHTP
atepo01p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~REG LDAAARGQTP
atepo05p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~P
atsora1x    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atfkb01p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atfkb09p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atrap03p    SAKTQPALTE HEDRLRAYLA ASPGADTRAV ASTLAVTRSV FEHRAVLLGD
atrap06p    ~~~TQPALTE HEDRLRAYLA ASPGVDTRAV ASTLAVTRSV FEHRAVLLGD
atrap04p    SAKTQPALTE HEDRLRAYLA ASPGADTRAV ASTLAVTRSV FEHRAVLLGD
atrap13p    SAKTQPALTE HEDRLRAYLA ASPGADIRAV ASTLAVTRSV FEHRAVLLGD
atrap01p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~LAVTRSL FEHRAVLLGD
atrap07p    SAKTLPALTE HEDRLRAYLA ASPGADMRAV GSTLALTRSV FEHRAVLLGH
atrap10p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~AV ASTLAVTRSV FEHRAVLLGD
atfkb04x    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atty104p    FGYRAVVLAR GEAELAGRLR ALAGGDPDAG VVTGAVVD.. ..........
atty106p    FGYRAVVLAR GEAELAGRLR ALAGGDPDAG VVTGAVVD.. ..........
atty101p    FGYRAVVLAR GEAELAGRLR ALAGGDPDAG VVTGAVVD.. ..........
atty102p    ~~~~~~~~~~ ~~~~~~~RLR ALAGGDPDAG VVTGAVVD.. ..........
atty100p    FGYRAVVLAR GEAELAGRLR ALAGGDPDAG VVTGAVLDGG VVVGAAPGGA
atnid05b    ~~~~~~~~~~ ~~~~~LLSTR ARFPRRAAVV GESMTELAEA LDAVAEGGPH
atty105b    LLEHPDEHPA DVGYTLITGR AHFGHRAAVI GESREELLDA LKALAEGREH
atnid06x    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~RSVAEERPE
atdebs01p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~GLATGNAD
atmon02p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~GALAAGEAS
atmon10p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ LGALAAGEAS
atmon04p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~LAAGETP
atmon07p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~ALAAGEES
atmon11p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~ALAAGEAS
atmon12p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~LAAGEPS
atmon05b    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~SLAAGEAS
atmon01p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~EALAAGDAS
atdebs02p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~ADGAVV
atdebs06p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~RAVAEGVAA
atave01p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~G LGALAAGEPD
atave07p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~G LGALAAGEPD
atave06p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~QA LTALAAGEPH
atave09p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ LTALAAGEPH
atnys01p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys11p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~AVATDG
atrif05p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~TALARGESA
atrif07p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~G LGALARGEAA
atrif08p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~AG LAALARGESA
atrif10p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ADSAEEARAG LGALARGEDA
atrif03p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~QDG LQALARGENA
atrif06p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~SREEAVTN LEALARGEDP
atrif04p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~RALARGESA
atrif01p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~V VVAGSREEAV
atnys02p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~AVVV GERREDFLRG LAALSTGAST
atfkb02p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~GEEV
atave11p    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~LHA LDALAEGAPT
atdebs03p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~AATA
atnid04p    ~~~~~SLADS AGIGHGLAVG RAALPHRAVL LGDGAAPLDA LAALASGEVS
atdebs05p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~ADRRIA
atdebs04p   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~ALAEGRPS
```

Fig 2c

```
atave02a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atave05a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~QALQAL AAGEPHPAVI
atave04a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~QALQAL AAGEPHPAVI
atave08a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~QALQAL AAGEPHPAVI
atave03a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~QALQAL AAGEPHPAVI
atrap02a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~DT RAVASTLAMT RSVFEYRAVL
atrap11a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~AVASTLAMT RSMFEHRGVL
atrap08a  MVISARTQSA LTEHEGRLRA YLAASPGVDM RAVASTLAIT RSVFEHRAVL
atrap12a  ~~~~~~~~~~ LTEHEGRLRA YLAASPGVDM RAVASTLAMT RSVFEHRAVL
atrap05a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~ASTLAVT RSVFEHRAVL
atrap09a  LVISAKTQSA LAEYEGRLRA YLAASPGVDM RAVASTLAMT RSVFEHRAVI
atfkb03a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~HRAAL
atfkb07x  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~L
atfkb08x  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnid01a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~KHRA VITGRTRTEL HTKLHTLDAI
atnid03a  ~~~~~~~~TQA DPQDIAHALA TTRTHFKHRA VITGRTRTEL HTKLHTLDAI
atnid02a  ~~~~~~~~~~ ~~~~~~HALA TTCTHFKHRA VITGRTRTEL HTKLHTLDAI
atnid00a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ SSALAALAAG
atfkb10a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atrap14a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~DFLRA LSKLADGAPW
atmon06a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~TGEPHA
atmon08a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~AGEEHP
atmon09a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~GEEHP
atepo02a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ AALSAVAQGQ
atepo03x  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~A VAVTSREGLL AALSAVAQGQ
atepo08a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~VAAQGQ
atepo00a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~SREGLR AALDAAAQGQ
atepo04a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~LR GALDAAAQQK
atnid07a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~A AAHDALLAVA DGRPSDAVVT
atty107a  PRDIAFSLAA TRAAFDHRAV LIGSDGAELA AALDAL...A EGRDGPAVVR
atsor02a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atsorb1a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys09a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~AD DPAAAPAWIT
atnys12a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~S DGRPDPGLVQ
atnys16a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~PD.LPEVAR
atnys17a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ APDGITAAAR
atnys03a  ~~~~~~~~~~ ~~~~~~~~~~ ~~GYALADGR ATFEHRAVLL PDGTELA..H
atnys15a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ PDAHE.G..H
atnys07a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~IAA DEA.DAAAAT
atnys08a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~ALAALAS GVA.DPAVVS
atnys05a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ AVRALTALAA ADA.DLSAVV
atnys06a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ATRALSALAT TAASDPSALT
atnys04a  ~~~~~~~~~~ ~~~~~~~~HR AVVLGTDRAE ALRALTALAA GE.TDPAALT
atnys14a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~DG LRTGLTAVAE GTTAPHTAEH
atnys00a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~ADAVEHAR
atnys10a  ~~~~~~~~~~ ~~~~~~~~~~ ~~VVAQDRDQ LIASLGALAA DRPDPAVVEG
atnys18a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ EGGAVTEVAR
atnys13a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~LLA GPDGVREAAR
atave10a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~LHALDALA GGRPVPGVVE
atrif02a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~R AVVLASDRAQ LCADLAAFGS
atmon03a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~A LAAGRAHPAL
atave12a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~QALDALA EGRSADGLIE
atrif09a  ~~~~~~~~~~ ~~~~~~~~~~ ~~GRALLGDR AVVVAGTDED AVAGLRALAR
atmon00a  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~LAEG
atty103a  GAGQGAGPGT AEVAGALAHA RTAFRHRAVV LGGNRAELLA GLRELAEEEH
```

Fig 2d

```
          101                                                  150
 atave00x AGEVLGVADE ADGG..VVFV FPGQGPQWPG MGRELLDASD VFRESVRACE
atdebs00p ASVVRGVA.R PSAP..VVFV FPGQGAQWAG MAGELLGESR VFAAAMDACA
 atepo06p PGAVRGRASG GSAP.KVVFV FPGQGSQWVG MGRKLMAEEP VFRAALEGCD
 atepo07p PGAVRGRASG GSAP.KVVFV FPGQGSQWVG MGRKLMAEEP VFRAALEGCD
 atepo01p PGAVRGRCSP GNVP.KVVFV FPGQGSQWVG MGRQLLAEEP VFHAALSACD
 atepo05p PAAARGHAST GSAP.KVVFV FPGQGSQWLG MGQKLLSEEP VFRDALSACD
 atsora1x ~~~~~~~~~~ ~~~~~~~VFV FAGQGAQWFG MGRALLQREP VFRTTIEQCS
 atfkb01p ~~~~SAVAGV AVEGARTVFV FPGQGSQWVG MGRELMGASE VFAARMRECA
 atfkb09p ~~~~~~~~~~ ~~~~~~~VFV FPGQGSQWVG MGRELMGCSE VFAARMRECA
 atrap03p D..TV..TGT AVSDPRVVFV FPGQGWQWLG MGSALRDSSV VFAERMAECA
 atrap06p D..TV..TGT AVSDPRVVFV FPGQGWQWLG MGSALRDSSI VFAERMAECA
 atrap04p D..AV..TGT AVTDPRVVFV FPGQGWQWLG MGSALRDSSV VFAERMAECA
 atrap13p D..TV..TGT AVTDPRIVFV FPGQGWQWLG MGSALRDSSV VFAERMAECA
 atrap01p D..SVTGTGT AVSDPRVVFV FPGQGWQWLG MGSALRTSSM VFAERMAECA
 atrap07p DTVTVTGTGT AVSNPRVVFV FPGQGWQWLG MGSALRGSSV VFAERMAECA
 atrap10p ETV....TGT AVSDPRIVFV FPGQGWQWLG MGSALRDSSV VFAERMAECA
 atfkb04x ~~~~~VVTGT ALTAPRTVFV FPGQGSQWLG MGRELMAESP VFAARMRQCA
 atty104p .......PET GSGGGGVVLV FPGQGTQWVG MGAGLLGSSE VFAASMRECA
 atty106p .......PET GSGGGGVVLV FPGQGTQWVG MGAGLLGSSE VFAASMRECA
 atty101p .......PET GSGGGGVVLV FPGQGTQWVG MGAGLLGSSE VFAASMRECA
 atty102p .......PET GSGGGGVVLV FPGQGTQWVG MGAGLLGSSE VFAASMRECA
 atty100p GAAGGAGAAG GAGGGGVVLV FPGQGTQWVG MGAGLLGSSE VFAASMRECA
 atnid05b ..PLAATGT. AGTADRVVFV FPGQGSQWAG MAEGLLERSG AFRSAADSCD
 atty105b HTVVRGDGT. AHPDRRVVFV FPGQGSQWPS MARDLLDRAP AFRETAKACD
 atnid06x PDVVL..GE. AGSDRAPAFV FPGQGAQWAG LGARLLADSP VFRARAEACA
atdebs01p GAAV...GT. SRAQQRAVFV FPGQGWQWAG MAVDLLDTSP VFAAALRECA
 atmon02p AGVVAG.VAG DVGPGP.VLV FPGQGAQWVG MGAQLLDESA VFAARIAECE
 atmon10p AGVVAG.VAG DVGPGP.VLV FPGQGSQWVG MGAQLLDESP VFAARIAECE
 atmon04p TDVVSG.AAA SSGAGP.VLV FPGQGSQWVG MGAQLLDESP VFAARIAECE
 atmon07p ADVVAG.VAG DVGPGP.VLV FPGQGSQWVG MGAQLLDESP VFAARIAECE
 atmon11p ADVVAG.VAG DVGPGP.VLV FPGQGSQWVG MGAQLLDESP VFAARIAECE
 atmon12p PDVVEGAVQG ASGAGP.VLV FPGQGSQWVG MGAQLLDESP VFAARIAECE
 atmon05b PDVVSGAV.G PTGPGP.VMV FPGQGGQWVG MGARLLDESP VFAARIAECE
 atmon01p PDVVCG.VAG DVGPGP.VLV FPGQGSQWVG MGAQLLGESA VFAARIDACE
atdebs02p PGVVTGSASD ....GGSVFV FPGQGAQWEG MARELL.PVP VFAESIAECD
atdebs06p PGATTGTASA ....GGVVFV FPGQGAQWEG MARGLL.SVP VFAESIAECD
 atave01p RRVTTGHAPG GDRGG.VVFV FPGQGGQWAG MGVRLLASSP VFARRMQACE
 atave07p RRVTTGHAPG GDRGG.VVFV FPGQGGQWAG MGVRLLASSP VFARRMQACE
 atave06p PHITTGHTRG GDRGG.VVFV FPGQGGQWAG MGLTLLTSSP VFAEHIDACE
 atave09p PHITTGHTRG SDRGG.VVFV FPGQGGQWAG MGLTLLTSSP VFAEHIDACE
 atnys01p ~~~~~~~~L. ADVEGRTVFV FPGQGSQWVG MGAQLLDESA VFAERIAECA
 atnys11p PSPVVARGV. ADVEGRTVFV FPGQGSQWVG MGSQLLDESA VFAERIAECA
 atrif05p SGLVTGT... AGMPGKTVWV FPGQGTQWAG MGRELLEASP VFAERIEECA
 atrif07p PGVVTGT... AGKPGKVVWV FPGQGTQWVG MGRELLDASP VFAERIKECA
 atrif08p ADVVTGTVAA SGVPGKLVWV FPGQGSQWVG MGRELLEASP VFAARIAECA
 atrif10p PGLVRGRVPA SGLPGKLVWV FPGQGTQWVG MGRELLEESP VFAERIAECA
 atrif03p PGVVTGT... AGKPGKVVWV FPGQGSQWMG MGRDLLDSSP VFAARIKECA
 atrif06p AAVVTGR... AGSPGKLVWV FPGQGSQWIG MGRELLDSSP VFAERVAECA
 atrif04p PGLLSGR..G SGVPGKVVWV FPGQGTQWAG MGRELLDSSE VFAARIAECE
 atrif01p TGLRALNTAG SGTPGKVVWV FPGQGTQWAG MGRELLAESP VFAERIAECA
 atnys02p AGLVSG..IA GPDPEGAVFV FPGQGSQWWG MGRELLATSE VFRTAIDDCA
 atfkb02p PGVVRGTADV TDT..RAVFV FPGQGSQWDG MGAELLATEP VFARRLGECA
 atave11p AGVVQGVAGP AA.DGKIAML FGGQGTHWEG MAQELLGSSP VFAQQMSDCA
atdebs03p DAVVEGV.TE VD.GRNVVFL FPGQGSQWAG MGAELLSSSP VFAGKIRACD
 atnid04p PDVVTG..SA AD.VRRVAFV FPGQGAQWAG MGAELLDSSP VFAAELARCE
atdebs05p DRTATGQ.GP NS.PRRVAMV FPGQGAQWQG MARDLLRESQ VFADSIRDCE
atdebs04p ADAVAPVTSA ...PRKPVLV FPGQGAQWVG MARDLLESSE VFAESMSRCA
```

Fig 2e

```
atave02a  HSSAPGGTGT GEAAGKTAFI CSGQGTQRPG MAHGLYHTHP VFAAALNDIC
atave05a  HSSAPGGTGT GEAAGKTAFI CSGQGTQRPG MAHGLYHTHP VFAAALNDIC
atave04a  HSSAPGGTGT GEAAGKTAFI CSGQGTQRPG MAHGLYHTHP VFAAALNDIC
atave08a  HSSAPGGTGT GEAAGKTAFI CSGQGTQRPG MAHGLYHTHP VFAAALNDIC
atave03a  HSSAPGGTGT GEAAGKTAFI CSGQGTQRPG MAHGLYHTHP VFAAALNDIC
atrap02a  IGDDTVTG.T AATDPRVVFV FPGQGSQRAG MGEELAAAFP VFARIHQQVW
atrap11a  LGDGTVSG.T AVSDPRVVFV FPGQGSQRAG MGEELAAAFP VFARIHQQVW
atrap08a  LGDDTVTG.T AATDPRVVFV FPGQGSQRAG MGEELAAAFP VFARIHQQVW
atrap12a  LGDDTVTG.T AVSDPRAVFV FPGQGSQRAG MGEELAAAFP VFARIHQQVW
atrap05a  LGDDTVTG.T TVSDPRVVFV FPGQGSQRAG MGEELAAAFP VFARIHQQVW
atrap09a  VGDDTVSG.T AATDPRVVFV FPGQGSQRAG MGAELAAAFP VFARIHQQVW
atfkb03a  IGTDLITG.T AEPDRRLVWL FSGQGSQRPG MGDELAAAYD VFARTRRDVL
atfkb07x  LGDTLITADP NAGSGPVVFV YSGQSTLHPH TGHQLAATYS VFADAWGEVL
atfkb08x  ~~~~~IGAPP ADQADELVFV YSGQGTQHPA MGEQLAAAFP VFADAWHDAL
atnid01a  Q.....GT AHPHPRLTLL FTGQGAQHRG MGQELYATDP HFAAALDEVC
atnid03a  Q.....GT AHPHPRLTLL FTGQGAQHPG MGQELYTTDP HFAAALDEVC
atnid02a  Q.....GT AHPHPRLTLL FTGQGAQHPG MGQELYTTDP HFAAALDEIC
atnid00a  QTPRGVRIGS TDADGRLALL FTGQGAQHPG MGQELYTTDP HFAAALDEVC
atfkb10a  ~~~EAPESSA EPPRSARRFL FDGQGAQRVG MGRELHGRFP VFAAAWDEVS
atrap14a  PGLTTATATA KARRVA..FL FDGQGTQRLG MGKELYDSYP AFARAWDTVS
atmon06a  ALVGPACSQA RVGGDDVVWL FSGQGSQLVG MGAGLYERFP VFAAAFDEVC
atmon08a  AVTRSREDGV AASG.AVVWL FSGQGSQLVG MGAGLYERFP VFAAAFDEVC
atmon09a  AVTRSREEAA VAASGDVVWL FSGQGSQLVG MGAGLYERFP VFAAAFDEVC
atepo02a  TPAGAARCIA SSSRGKLAFL FTGQGAQTPG MGRGLCAAWP AFREAFDRCV
atepo03x  TPPGAARCIA SSSRGKLAFL FTGQGAQTPG MGRGLCAAWP AFREAFDRCV
atepo08a  TPAGAARGRA ASSPGKLAFL FAGQGAQVPG MGRGLWEAWP AFRETFDRCV
atepo00a  TSPGAVRSIA DSSRGKLAFL FTGQGAQTLG MGRGLYDVWS AFREAFDLCV
atepo04a  TPQGAVRGKA VSSRGKLAFL FTGQGAQMPG MGRGLYETWP AFREAFDRCV
atnid07a  GIAR..... ..RGRDVAFL FSGQGAQRAG AGRELYASFP VFAQALDEVA
atty107a  GVRD..... ..RDGRMAFL FTGQGSQRAG MAHDLHAAHT FFASALDEVT
atsor02a  ~~~~~~~~~~ ~~~~~~~AVL FTGQGSQRPT MGRALYDAFP VFRDALDTVA
atsorb1a  ~~~~~~~~~~ ~~~~~~~AIL FTGQGSQRPT MGRALYDAFP VFRGALDAAA
atnys09a  GTT.R..... ..AETRLAVL FTGQGAQRLG AGRELAARFP AFATALDAAL
atnys12a  GTA..... ..GRGRTAFL FTGQGSQRPG MGRELHDRYP VFADALDEVL
atnys16a  GAA.TPH... ..RT...AFL FSGQGAQRSG MGRELHAAFP VFAAAFDEVV
atnys17a  AEA.RER... ..ST...AFL FSGQGAQRSG MGRELHAAFP VFAAAFDEVV
atnys03a  GTA.GEG... ..PC...AVL FSGQGSQRPG MGRELHARFP VFAAAFDEIT
atnys15a  .AA.GRT... ..RC...AAL FSGQGAQRLG MGRELHARFP VFARALDTAV
atnys07a  GRV.GAG... ..RH...AVL FSGQGAQRLG MGRELYERFP VFAEALDVVV
atnys08a  DAV.STG... ..GS...AVL FTGQGAQRLG MGRELYGRFP VFAEALDVVV
atnys05a  GDT.RTG... ..RH...AVL FSGQGSQRLG MGRELYERFP VFAEALDVAI
atnys06a  GTV.TMG... ..RC...AVL FSGQGSQRLG MGRELYERFP VFAEALDVVI
atnys04a  GTV.RTG... ..RT...AFL FSGQGSQRLG MGRVLYERFP AFAEALDTVL
atnys14a  HLQ.GTG... ..KR...AVL FSGQGSQRLG MGRELHERHP VFAEAFDSVL
atnys00a  GAA.HQR... ..RT...AVL FSGQGSQRPG MGRELAARFP VFADALDDAL
atnys10a  EAA.GRG... ..RT...AVL FTGQGSQRAA MGRELHEVQP EFAAAFDAVC
atnys18a  GAV.PTG... ..DRGGLAVL FSGQGSQRPG MGRELHARYP VFAAAFDETV
atnys13a  AAA.PRT... ..P.GRTAFL FSGQGAQHAL MGHDLYQRFP VYADALDTVL
atave10a  GRT.TSG... .....ELAVL FAGQGTQRAG MGRELYEAYP VFAQAIDEIC
atrif02a  GVVTGTP... ..VDGKLAVL FTGQGSQWAG MGRELAETFP VFRDAFEAAC
atmon03a  TRAAGPA... ..RNGGTAFL FTGQGSQRPG MGRQLYDTFD VFAESLDETC
atave12a  GSVGPRGGHS GRRRGKTAML FAGQGTQRVG MGRQLYAAHP AYADALDQVL
atrif09a  GDRAPGVLTG SAKHGKVVYV FPGQGSQRLG MGRELYDRYP VFATAFDEAC
atmon00a  AETASIVRGE AYTEGRTAFL FSGQGAQRLG MGRELYAVFP VFADALDEAF
atty103a  PGPRVVTGTA PATERRTAFL FSGQGSQRAG SGRGLYRRHP VFARALDEVC
                                ***                          GQG
```

Fig 2f

```
          151                                                       200
 atave00x AAFAPYVDWS VEQVLRDSPD A..... .....PG LDRVDVVQPT
atdebs00p RAFEPVTDWT LAQVL.DSPE Q..... .....S. .RRVEVVQPA
 atepo06p RAIEAEAGWS LLGEL..... .SA..... ....DEAASQ LGRIDVVQPV
 atepo07p RAIEAEAGWS LLGEL..... .SA..... ....DEAASQ LGRIDVVQPV
 atepo01p RAIQAEAGWS LLAEL..... .AA..... ....DEGSSQ LERIDVVQPV
 atepo05p RAIQAEAGWS LLAEL..... .AA..... ....DETTSQ LGRIDVVQPA
 atsora1x SFIQQNLGWS LLDEL..... .MT..... ....DRESSR LDEIDVSLPA
 atfkb01p AVLEPHTGWD LLDVL..... ..... .....GEAVV VDRVEVLQPA
 atfkb09p AVLEPYTGWD LLDVL..... ..... .....GEAVV AERVEVLQPA
 atrap03p AALSEFVDWD L.TVL..... ..... .....DDPAV VDRVDVVQPA
 atrap06p PALREFVDWD LFTVL..... ..... .....DDPAV VDRVDVVQPA
 atrap04p AALSEFVDWD LFAVL..... ..... .....DDPAV VDRVDVVQPA
 atrap13p AALREFVDWD LFTVL..... ..... .....DDPAV VDRVDVVQPA
 atrap01p AALSEFVDWD LFAVL..... ..... .....DDPAV VARVDVVQPA
 atrap07p AALSEFVDWD LFAVL..... ..... .....DDPAV VDRVDVVQPA
 atrap10p AALSEFVDWD LFAVL..... ..... .....DDPAV VDRVDVVQPA
 atfkb04x DALAEHTGRD LIAML..... ..... .....DDPAV KSRVDVVHPV
 atty104p RALSVHVGWD LLEVVSG... ..... .......GAG LERVDVVQPV
 atty106p RALSVHVGWD LLEVVSG... ..... .......GAG LERVDVVQPV
 atty101p RALSVHVEWD LLEVVSG... ..... .......GAG LERVDVVQPV
 atty102p RALSVHVEWD LLEVVSG... ..... .......GAG LERVDVVQPV
 atty100p RALSVHVGWD LLEVVSG... ..... .......GAG LERVDVVQPV
 atnid05b AALRPYLGWS VLSVLRGEPD ..... .......APS LDRVDVVQPV
 atty105b AALSVHLDWS VLDVLQEKPD ..... .......APP LSRVDVVQPV
 atnid06x RALEPHLDWS VLDVLAGAPG ..... .......TPP IDRADVVQPV
atdebs01p DALEPHLDFE VIPFLRAEAA RRE..... ....QDAALS TERVDVVQPV
 atmon02p RALSAHVDWS LSAVLRG..D ..... .......GSE LSRVEVVQPV
 atmon10p RALSAYVDWS LSAVLRG..D ..... .......GSE LSRVEVVQPV
 atmon04p QALSAYVDWS LSDVLRG..D ..... .......GSE LSRVEVVQPV
 atmon07p QALSAYVDWS LSAVLRG..D ..... .......GSE LSRVEVVQPV
 atmon11p QALSAHVDWS LSDVLRG..D ..... .......GSE LSRVEVVQPV
 atmon12p RALSAHVDWS LSAVLRG..D ..... .......GSE LSRVEVVQPV
 atmon05b QALSAYVDWS LTDVLRG..D ..... .......GSE LARIDVVQPV
 atmon01p QALSPYVDWS LTEVLRG..D ..... .......GRE LSRVDVVQPV
atdebs02p AVLSEVAGFS VSEVLEPRPD ..... .......APS LERVDVVQPV
atdebs06p AVLSEVAGFS ASEVLEQRPD ..... .......APS LERVDVVQPV
 atave01p EALAPWVDWS VVDILRRDAG ..... .......DAV WERADVVQPV
 atave07p EALAPWVDWS VVDILRRDAG ..... .......DAV WERADVVQPV
 atave06p KALTPWVPWS LTDILHRDPD ..... .......DPA WQQADVVQPV
 atave09p KALTPWVPWS LTDILHRDPD ..... .......DPA WQQADVVQPV
 atnys01p AALAEFTDWS LVDVLRGVVG ..... .......APS LERVDVVQPA
 atnys11p AALAEFTDWS LVDVLRGVVG ..... .......APS LERVDVVQPA
 atrif05p AALQPWIDWS LLDVLRG..E ..... .......GE. LDRVDVLQPA
 atrif07p AALDQWTDWS LLDVLRG..D ..... .......GD. LDSVEVLQPA
 atrif08p AALEPWIDWS LLDVLRG..E ..... .......GD. LDRVDVVQPA
 atrif10p AALEPWIGWS LFDVLRG..D ..... .......GD. LDRVDVLQPA
 atrif03p AALEQWTDWS LLDVLRG..D ..... .......ADL LDRVDVVQPA
 atrif06p AALEPWIDWS LLDVLRG..E ..... .......SDL LDRVDVVQPA
 atrif04p TALGRWVDWS LTDVLRG..E ..... .......ADL LDRVDVVQPA
 atrif01p AALAPWIDWS LVDVLRG..E ..... .......GD. LGRVDVLQPA
 atnys02p TALAPYVDWS LHDVLAGEGD ..... .......PAL LERVDVVQPA
 atfkb02p EALAPYTGWD LLDVIARRPG ..... .......APE LDRVDVVQPA
 atave11p QALEPYLDWS LLDVLRGAPD ..... .......APP LQRVDVVQPV
atdebs03p ESMAPMQDWK VSDVLRQAPG ..... .......APG LDRVDVVQPV
 atnid04p AALEPFVDWS LTDVLRGAPG ..... .......APG LDRVDVVQPV
atdebs05p RALAPHVDWS LTDLL...SG ..... .......ARP LDRVDVVQPA
atdebs04p EALSPHTDWK LLDVVRGDGG ..... .......PDP HERVDVLQPV
```

Fig 2g

```
atave02a  THLDPHLDHP LLPLLTQ..N DNDN..... ....EDAAAL LQQTRYAQPA
atave05a  THLDPHLDHP LLPLLTQNDN DNDN..... ....EDAAAL LQQTPYAQPA
atave04a  THLDPHLDHP LLPLLTQDPN TQDT..... .TTLEEAAAL LQQTPYAQPA
atave08a  THLDPHLDHP LLPLLTQDPN TQDT..... .TTLEEAAAL LQQTPYAQPA
atave03a  THLDPHLDHP LLPLLTQDPN TQDT..... .TTLEEAAAL LQQTRYAQPA
atrap02a  DLLDVP.DLD .......... .......... .......... VNETGYAQPA
atrap11a  DLLDVP.DLD .......... .......... .......... VNETGYAQPA
atrap08a  DLLDVP.DLE .......... .......... .......... VNETGYAQPA
atrap12a  DLLDVP.DLE .......... .......... .......... VNETGYAQPA
atrap05a  GLLDVP.DLE .......... .......... .......... VNETGYAQPA
atrap09a  DLLDVP.DLE .......... .......... .......... VNETGYAQPA
atfkb03a  DALQVPAGLD .......... .......... .......... VHDTGYAQPA
atfkb07x  GHLN..ADQG .......... .......... .......... P.......AT
atfkb08x  RRLD...DPD .......... .......... .......... PHDPTRSQHT
atnid01a  EELQR..... .........C GTQNLREVMF TPD...QPDL LDRTEYTQPA
atnid03a  EELQR..... .........C GTQNLREVMF TPD...QPDL LDRTEYTQPA
atnid02a  EELQR..... .........C GTQNLREVMF TPD...QPDL LDRTEYTQPA
atnid00a  EELQR..... .........C GTQNLREVMF TPD...QPDL LDRTEYTQPA
atfkb10a  DAFGKHLE.. .......... ..HSPTDVFH GEHGD....L AHDTLYAQVG
atrap14a  AGFDKHLD.. .......... ..HSLTDVCF GEGGSTTAGL VDDTLYAQAG
atmon06a  GLLEGPL... .........GV EAGGLREVVF RGPR....ER LDHTVWAQAG
atmon08a  GLLEGPL... .........GV EAGGLREVVF RGPR....ER LDHTMWAQAG
atmon09a  GLLEGEL... .........GV GSGGLREVVF WGPR....ER LDHTVWAQAG
atepo02a  ALFDRELDRP .......... ...LREVMW AEAGSAESLL LDQTAFTQPA
atepo03x  ALFDRELDRP .......... ....LREVMW AEPGSAESLL LDQTAFTQPA
atepo08a  TLFDRELHQP .......... ....LCEVMW AEPGSSRSSL LDQTAFTQPA
atepo00a  RLFNQELDRP .......... ....LREVMW AEPASVDAAL LDQTAFTQPA
atepo04a  ALFDREIDQP .......... ....LREVMW AAPGLAQAAR LDQTAYAQPA
atnid07a  GGFDAHLERP .......... ....LLQVMF AEPGTADAAL LDRTAYAQPA
atty107a  DRLDPLLGRP .......... ....LGALLD ARPGSPEAAL LDRTEYTQPA
atsor02a  AHLDRDLDRP .......... ....LRDVLF APDGSEQAAR LDQTAFTQPA
atsorb1a  AHLDRDLDRP .......... ....LRDVLF APDGSEQAAR LDQTAFTQPA
atnys09a  DAFTPHLDRP .......... ....LREVLW ....GTDAAL LDRTAYAQPA
atnys12a  ARLDDGPDRP .......... ....LREVLF AAPDSAEAAL LDRTGYAQPA
atnys16a  AVLDAELGSD .........AD GGVSLREVMW GGG....SEL LDRTRFTQPA
atnys17a  AVLDAELATG .........SG GGVSLREVMW GGG....SEL LDRTRFTQPA
atnys03a  ALLDTHLDRP .......... ....LREVVW GTD....ADL LNDTGWAQPA
atnys15a  DLLDAELGGT .......... ....LREVIW GTD....DAP LNETGFTQPA
atnys07a  DHLDAALPAQ .........AG ....LREVMW GDD....AEL LNETGWTQPA
atnys08a  DHLDAALPAQ .........AG ....LREVMW GDD....VEL LNETGWTQPA
atnys05a  DHLDAALPAQ .........AS ....LREVMW GDD....VEL LDETGWTQPA
atnys06a  DHLDAALPAQ .........AG ....LREVMW GDD....VEL LNETGWTQPA
atnys04a  TALDAELGHP .......... ....LRDIIW GED....AQL VDRTGYTQPA
atnys14a  ARLDDRLDTP .......... ....LRDVVW GTD....EEA LHATGNTQPA
atnys00a  RALDRHLDGP .......... ....VREVMW GTD....AAL LDRTGWTQPA
atnys10a  AVFDPLLDRP .......... ....LREVVF AEDGSDEAAL LDETGWTQPA
atnys18a  ALLDARL... .......... .GTSLRDIVW DQDRTR.... LDDTRHTQPA
atnys13a  AQFDTVLDVP .......... ....LRAALF AAPGTPEAAL LDQTGFTQPA
atave10a  AEADTARTDP .........GA PG..LRDVLF APQDSPEGRL IEDTGFAQPA
atrif02a  EAVDTHL... .......... RERPLREVVF .....DDSAL LDQTMYTQGA
atmon03a  ARLDPLLEQP .......... ....LKPVLF APADTAQAAV LHGTGMTQAA
atave12a  AELDGHLDQP .........LR PLIHASADL. .ADVADAADV LDRTRYAQPA
atrif09a  EQLDVCL..A .........GR AGHRVRDVVL GE.VPAETGL LNQTVFTQAG
atmon00a  AALDVHLDRP LREIVLGETD SGGNVSGENV IGEGADHQAL LDQTAYTQPA
atty103a  AALEPHLHRP .......... ....LRDLMF AEPGSPEAEP LDRTEFTQPA
```

Fig 2h

```
          201                                                  250
 atave00x LFAVMISLAA L.WRSQGVEP CAVLGHSLGE IAAAHVSGGL SLADAARVVT
atdebs00p LFAVQTSLAA L.WRSFGVTP DAVVGHSIGE LAAAHVCGAA GAADAARAAA
 atepo06p LFAMEVALSA L.WRSWGVEP EAVVGHSMGE VAAAHVAGAL SLEDAVAIIC
 atepo07p LFAMEVALSA L.WRSWGVEP EAVVGHSMGE VAAAHVAGAL SLEDAVAIIC
 atepo01p LFALAVAFAA L.WRSWGVAP DVVIGHSMGE VAAAHVAGAL SLEDAVAIIC
 atepo05p LFAIEVALSA L.WRSWGVEP DAVVGHSMGE VAAAHVAGAL SLEDAVAIIC
 atsora1x IISIEIALAA Q.WRAWGVEP AFVVGHSTGE IAAAHVAGVL SIEDAMRTIC
 atfkb01p SWAVAVSLAA L.WQAHGVVP DAVVGHSQGE IAAACVAGAL SLEDAARVVA
 atfkb09p SWAVAVSLAA L.WQAHGVSP DAVIGHSQGE IAAACVAGAL SLEDAARIVA
 atrap03p SWAVMVSLAA V.WQAAGVRP DAVIGHSQGE IAAACVAGAV SLRDAARIVT
 atrap06p SWRMMVSLAA V.WQAAGVRP DAVIGHSQGE IAAACVAGAV SMRDAARIVT
 atrap04p SWAVMVSLAA V.WQAAGVRP DAVIGHSQGE IAAACVAGAV SLRDAARIVT
 atrap13p SWAMMVSLAA V.WQAAGVRP DAVIGHSQGE IAAACVAGAV SLRDAARIVT
 atrap01p SWAVMVSLAA V.WQAAGVRP DAVVGHSQGE IAAACVAGAV SLRDAARVVT
 atrap07p SWAVMVSLAA V.WQADGVRP DAVIGHSQGE IAAACVAGAV SLRDAARSVT
 atrap10p SWAVMVSLAA V.WQAAGVRP DAVIGHSQGE IAAACVAGAV SMRDAARIVT
 atfkb04x CWAVMVSLAA V.WEAAGVRP DAVVGHSQGE IAAACVAGAI SLEDGARLVA
 atty104p TWAVMVSLAR Y.WQAMGVDV AAVVGHSQGE IAAATVAGAL SLEDAAAVVA
 atty106p TWAVMVSLAR Y.WQAMGVDV AAVVGHSQGE IAAATVAGAL SLEDAAAVVA
 atty101p TWAVMVSLAR Y.WQAMGVDV AAVVGHSQGE IAAATVAGAL SLEDAAAVVA
 atty102p TWAVMVSLAR Y.WQAMGVDV AAVVGHSQGE IAAATVAGAL SLEDAAAVVA
 atty100p TWAVMVSLAR Y.WQAMGVDV AAVVGHSQGE IAAATVAGAL SLEDAAAVVA
 atnid05b LFTMMVSLAA V.WRALGVEP AAVVGHSQGE IAAAHVAGAL SLDDSARIVA
 atty105b LFTMMLSLAA C.WRDLGVHP AAVVGHSQGE IAAACVAGAL SLEDAARIVA
 atnid06x LFTTMVSLAA L.WEAHGVRP AAVVGHSQGE VAAACVAGAL SLDDAALVIA
atdebs01p MFAVMVSLAS M.WRAHGVEP AAVIGHSQGE IAAACVAGAL SLDDAARVVA
 atmon02p LWAVMVSLAA V.WADYGVTP AAVIGHSQGE MAAACVAGAL SLEDAARIVA
 atmon10p LWAVMVSLAA V.WADYGVTP AAVIGHSQGE MAAACVAGAL SLEDAARIVA
 atmon04p LWAVMVSLAA V.WADYGVTP AAVVGHSQGE MAAACVAGAL SLEDAARIVA
 atmon07p LWAVMVSLAA V.WADYGVTP AAVIGHSQGE MAAACVAGAL SLEDAARVVA
 atmon11p LWAVMVSLAA V.WADYGITP AAVIGHSQGE MAAACVAGAL SLEDAARIVA
 atmon12p LWAVMVSLAS V.WADYGITP AAVIGHSQGE MAAACVAGAL SLEDAARIVA
 atmon05b LWAVMVALAA V.WADQGIEP AAVVGHSQGE IAAACVVGAI SLDEAARIVA
 atmon01p LWAVMVSLAA V.WADHGVTP AAVVGHSQGE IAAVVVAGAL TLEDGAKIVA
atdebs02p LFAVMVSLAR L.WRACGAVP SAVIGHSQGE IAAAVVAGAL SLEDGMRVVA
atdebs06p LFSVMVSLAR L.WGACGVSP SAVIGHSQGE IAAAVVAGVL SLEDGVRVVA
 atave01p LFSVMVSLAA L.WRSYGIEP DAVLGHSQGE IAAAHVCGAL SLKDAAKTVA
 atave07p LFSVMVSLAA L.WRSYGIEP DAVLGHSQGE IAAAHVCGAL SLKDAAKTVA
 atave06p LFSIMVSLAA L.WRSYGIEP DAVLGHSQGE IAAAHICGAL SLKDAAKTVA
 atave09p LFSIMVSLAA L.WRSYGIEP DAVLGHSQGE IAAAHICGAL SLKDAAKTVA
 atnys01p SFAVMVSLAA L.WGSRGVLP DAVVGHSQGE IAAAVVSGAL SLRDGARVVA
 atnys11p SFAVMVSLAA L.WRSRGVLP DAVVGHSQGE IAAAVVSGAL SLRDGARVVA
 atrif05p CFAVMVGLAA V.WASVGVVP DAVLGHSQGE IAAACVSGAL SLEDAAKVVA
 atrif07p CFAVMVGLAA V.WESAGVRP DAVVGHSQGE IAAACVSGAL TLDDAAKVVA
 atrif08p SFAVMVGLAA V.WSSVGVVP DAVLGHSQGE IAAACVSGAL SLQDAAKVVA
 atrif10p CFAVMVGLAA V.WSSAGVVP DAVLGHSQGE IAAACVSGAL SLEDAAKVVA
 atrif03p SFAMMVGLAA V.WTSLGVTP DAVLGHSQGE IAAACVSGAL SLDDAAKVVA
 atrif06p SFAMMVGLAA V.WQSVGVRP DAVVGHSQGE IAAACVSGAL SLQDAAKVVA
 atrif04p SFAVMVGLAA V.WASLGVEP EAVVGHSQGE IAAACVSGAL SLEDAAKVVA
 atrif01p CFAVMVGLAA V.WESVGVRP DAVVGHSQGE IAAACVSGAL SLEDAAKVVA
 atnys02p LFAMMVGLSA L.WRSHGVVP AAVVGHSQGE IAAACVAGAL SLADAARVVA
 atfkb02p SFAMMVALAE L.WRAHGVAP AAVVGHSQGE VAAACVAGVL TLDDAAKVVA
 atave11p LFAVMVSLAA L.WRSYGVHP DAVAGHSQGE IAAAYVAGAL SLDDAARVTA
atdebs03p LFAVMVSLAE L.WRSYGVEP AAVVGHSQGE IAAAHVAGAL TLEDAAKLVV
 atnid04p TFAVVVALAA M.WRWLGVEP AAVVGHSQGE IAAAHVAGVL SLEDAARVVA
atdebs05p LFAVMVSLAA L.WRSHGVEP AAVVGHSQGE IAAAHVAGAL TLEDAAKLVA
atdebs04p LFSIMVSLAE L.WRAHGVTP AAVVGHSQGE IAAAHVAGAL SLEAAAKVVA
```

Fig 2i

```
atave02a  QRATLMQTMP P..GTMTTLH TTPHHIT..H HLTAHE...N DLAIAAINTP
atave05a  QRATLMQTMP P..GTMTTLH TTPHHIT..H HLTAHE...N DLAIAAINTP
atave04a  QRATLMQTMP P..GTMTTLH TTPHHIT..H HLTAHE...N DLAIAAINTP
atave08a  QRATLMQTMP P..GTMTTLH TTPHHIT..H HITAHE...N DLAIAAINTP
atave03a  QRATLMQTMP P..GTMTTLH TTPHHIT..H HLTAHE...N DLAIAAINTP
atrap02a  ARARLMQALP AG.GVMAAVP VSEDEARAVL G........E GVEIAAVNGP
atrap11a  ARARLMQALP AG.GVMVAVP VSEDEARAVL G........E GVEIAAVNGP
atrap08a  ARARLMQALP AG.GVMVAVP VSEDEARAVL G........E GVEIAAVNGP
atrap12a  ARARLMQALP AG.GVMVAVP VSEDEARAVL G........E GVEIAAVNGP
atrap05a  ARARLMQALP PG.GVMVAVP VSEDEARAVL G........E GVEIAAVNGP
atrap09a  ARARLMQALP AG.GVMVAVP VSEDEARAVL G........E GVEIAAVNGP
atfkb03a  ARARLMQALP PG.GAMAAVS ASERDALPLL C........E GVEIAAVNGP
atfkb07x  ARSRLMDELP TG.GAMVTVL TSEENALRAL R........P GVEIAAVNGP
atfkb08x  TRARLMHTLP PP.GAMVTVL TSEEEARQAL R........P GVEIAAVFGP
atnid01a  ARAHLMGQLP HG.GAMLSVQ AAEHDLDQLA ....HT...H GVEIAAVNGP
atnid03a  ARAHLMGQLP HG.GAMLSVQ AAEHDLDQLA ....HT...H GVEIAAVNGP
atnid02a  ARAHLMGQLP HD.GAMLSVQ AAEHDLDQLA ....HT...H GVEIAAVNGP
atnid00a  ARAHVMGQLP HG.GAMLSVQ AAEHDLDQLA ....HT...H GVEIAAVNGP
atfkb10a  ARGRALRALP P..GAMTAVE GSPAEVG..A FTD.......  .LDIAAVNGP
atrap14a  ARGRALRTTP P..GAMVALR AGEEEVR..E FLSRTG...A ALDLAAVNSP
atmon06a  ARARLMGGLP EG.GAMCAVQ ATPAELAA..  ..DVDG...S AVSVAAVNTP
atmon08a  ARARLMGGLP EG.GAMCAVQ ATPAELAA..  ..DVDD...S GVSVAAVNTP
atmon09a  ARARLMGGLP EG.GAMCAVQ ATPAELAA..  ..DVDG...S SVSVAAVNTP
atepo02a  ARGRLMQGLS AG.GAMVSLG APEAEVA..A AVAPHA...A SVSIAAVNGP
atepo03x  ARGRLMQGLS AG.GAMVSLG APEAEVA..A AVAPHA...A SVSIAAVNGP
atepo08a  ARGRLMQALP AG.GAMVSIA APEADVA..A AVAPHA...A LVSIAAVNGP
atepo00a  ARGRLMQALP AG.GAMVSIE APEADVA..A AVAPHA...A SVSIAAVNAP
atepo04a  ARGRLMQALP AG.GAMVAIA ASEAEVA..A SVAPHA...A TVSIAAVNGP
atnid07a  ARGRLMQRLP EG.GAMVAVR ATEQEVAELE WIAGGR....  AV.VAAFNGP
atty107a  ARGRLMQRLP PG.GAMVSVR AGEDEVRAL:  .LAGRE...D AVCVAAVNGP
atsor02a  ARAKLMQALP QG.GAMVTLR ASEEEVRDL.  .LQPYD...G RASLAALNGP
atsorb1a  ARAKLMQALP QG.GAMVTLQ ASEQEARDL.  .LQAAE...G RVSLAAVNGH
atnys09a  ARGRLMQALP DG.GAMIAVQ ASEADVAPL.  .LAGHE...D QVAIAAVNGP
atnys12a  ARGRLMQALP EG.GAMVALE AAEDEVLPL.  .LEGLT...D RVSVAAVNGP
atnys16a  ARASLMDALP VG.GVMVAVE AAEAEVVPL.  .L...V...D GVAIAAVNGP
atnys17a  ARASLMDALP VG.GVMVAVE AAEAEVVPL.  .L...V...D GVAIAAVNGP
atnys03a  ARARLMQALP RG.GAMLAIR ATEDEVTPH.  .L...T...D DVSIAAVNGP
atnys15a  ARAGLMQALP RG.GAMVAVE ATEDEVSPL.  .L...T...D GVAIAAINGP
atnys07a  ARATLMQALP AG.GAMIAVQ ATEDEVTPH.  .L...T...D DVAIAAINGP
atnys08a  ARATLMQALP TG.GAMIAVQ ATEDEVTPH.  .L...T...D EVAIAAVNGP
atnys05a  ARATLMQALP TG.GAMIAIQ AAEDEVTQH.  .L...T...D DVSIAAVNGP
atnys06a  ARATLMQALP AG.GAMIAVQ ATEDEVIPH.  .L...T...D EVAIAAVNGP
atnys04a  ARAVLMQSLP EG.GAMIAVQ ATEDEVLPL.  .L...T...D DVSIAAVNSP
atnys14a  ARAALMQRLP AG.GAMIAVE ATEDEVTPL.  .L...T...D GVSLAAVNGP
atnys00a  ARATLMQALP AG.GAMAALE ATEDEVAPL.  .L...G...A HLALAAVNGP
atnys10a  ARATLMQALP TG.GAMIAIQ ATEDEIAAH.  .L...D...D TVAIAAVNGP
atnys18a  ARATAMSELP PG.GAMVALE ATEDEVRPL.  .L...T...D DLAIAAVNAP
atnys13a  ARASLMQQLP RD.GAMVALE ATEDEVAPL.  .L...T...D GVALAAVNGP
atave10a  ARGRLMQGLP SG.GAMVAIE ASEDEILPL.  .PDEYA...S RVAHAAVNGP
atrif02a  ARGRLMQALP AG.GAMVAVQ ATEDEVAPL.  .LDGT.....  .VCVAAVNGP
atmon03a  ARGRLMQALP AG.GAMLAVQ AAEDDVLPL.  .LAGQE...E RLSLAAVNGP
atave12a  ARGRLMEQLA PG.GAMVAVR ASEAEAR..Q ALDGRE...A RVSVAAVNGP
atrif09a  ARGRLMQALA PG.GAMVAVA ASEAEVAELL G........D GVELAAVNGP
atmon00a  TRGRLMQAVR AP.GAMAAWQ ATADEAA..E QLAGHE...R HVTVAAVNGP
atty103a  ARGRLMQALP AG.GAMAALR ATAEEIAPL.  .LERRA...G ELALAAVNGP
           *                                                  Arginine
```

Fig 2l

	301				350	
atave00x	RSTVVSGARE	AVADLVADLT	AAQVRTRMIP	.VDVPAHSPL	MYAIEERVV.	Load AT
atdebs00p	RSVLLTGSPE	PVARRVQELS	AEGVRAQVIN	.VSMAAHSAQ	VDDIAEGMR.	Load AT
atepo06p	RSTVLAGEPA	ALSEVLAALT	AKGVFWRQV.	KVDVASHSPQ	VDPLREEL.I	
atepo07p	RSTVLAGEPA	ALSEVLAALT	AKGVFWRQV.	KVDVASHSPQ	VDPLREEL.I	
atepo01p	RSTVLSGEPA	AIGEVLSSLN	AKGVFCRRV.	KVDVASHSPQ	VDPLREDL.L	
atepo05p	RSTVLAGEPA	ALAEVLAILA	AKGVFCRRV.	KVDVASHSPQ	IDPLRDEL.L	
atsora1x	DSTVLAGEPD	ALDALLQALE	RKNVFCRRV.	AMDVAPHCPQ	VDCLRDEL.F	Benzoate-CoA
atfkb01p	ESTVVAGDPA	AVERVLARYE	AEGVRVRRI.	AVDYASHTPH	VEAIEAQL.A	
atfkb09p	ESTVVAGDPS	AVERVLARYE	AEGVRVRRI.	AVDYASHTPH	VEAIQEQL.A	
atrap03p	ASTVIAGTPE	AVDHVLTAHE	ARGVRVRRI.	TVDYASHTPH	VELIRDEL.L	
atrap06p	ASTVIAGTPE	AVDHVLTAHE	ARGVRVRRI.	TVDYASHTPH	VELIRDEL.L	
atrap04p	ASTVIAGTPE	AVDHVLTAHE	ARGVRVRRI.	TVDYASHTPH	VELIRDEL.L	
atrap13p	ASTVIAGTPE	AVDHVLTAHE	AQGVRVRRI.	TVDYASHTPH	VELIRDEL.L	
atrap01p	ASTVVAGAPE	AVDRVLAVHE	ARGVRVRRI.	AVDYASHTPH	VELIRDEL.L	
atrap07p	ASTVVAGAPE	AVDRVLAVHE	ARGVRVRRI.	AVDYASHTPH	VELIRDEL.L	
atrap10p	ASTVIAGTPE	AVDHVLTALR	QRGAGAAD..	HVDYASHTPH	VELIRDEL.L	
atfkb04x	ATTIVSGRPD	AVETLIADYE	ARGVWVTRL.	VVDCPTHTPF	VDPLYDEL.Q	C5 unit
atty104p	ASTVVSGDRR	AVAGYVAVCQ	AEGVQARLIP	.VDYASHSRH	VEDLKGELE.	
atty106p	ASTVVSGDRR	AVAGYVAVCQ	AEGVQARLIP	.VDYASHSRH	VEDLKGELE.	
atty101p	ASTVVSGDRR	AVAGYVAVCQ	AEGVQARLIP	.VDYASHSRH	VEDLKGELE.	
atty102p	ASTVVSGDRR	AVAGYVAVCQ	AEGVQARLIP	.VDYASHSRH	VEDLKGELE.	
atty100p	ASTVVSGDRR	AVAGYVAVCQ	AEGVQARLIP	.VDYASHSRH	VEDLKGELE.	
atnid05b	GSCAVAGDPE	ALAELVALLT	GEGVHARPIP	GVDTAGHSPQ	VDALRAHL.L	Etmalonyl-CoA
atty105b	GTAAVAGDVD	ALRELLAELT	AEGIRAKPIP	GVDTAGHSAQ	VDGLKEHL.F	Etmalonyl-CoA
atnid06x	ASVTVSGDAL	ALEEFGARLS	AEGVLRWPLP	GVDFAGHSPQ	VEEFRAEL.L	MeOmalonyl CoA
atdebs01p	RSVVVAGDSD	ELDRLVASCT	TECIRAKRL.	AVDYASHSSH	VETIRDALHA	
atmon02p	SSTVISGPPE	HVAAVVADAE	ARGLRARVID	.VGYASHGPQ	IDQLHDLL.T	
atmon10p	SSTVISGPPE	HVAAVVADAE	ARGLRARVID	.VGYASHGPQ	IDQLHDLL.T	
atmon04p	SSTVISGPPE	HVAAVVAEAE	ARGLRARVID	.VGYASHGPQ	IDQLHDLL.T	
atmon07p	SSTVISGPPE	HVAAVVADAE	ERGLRARVID	.VGYASHGPQ	IDQLHDLL.T	
atmon11p	SSTVISGPPE	HVAAVVADAE	AQGLRARVID	.VRYASHGPQ	IDQLHDLL.T	
atmon12p	SSTVISGPPE	HVAAVVADAE	ARGLRARVID	.VGYASHGPQ	IDQLHDLL.T	
atmon05b	SSTVISGPPE	GIAAVVADAQ	ERGLRARAVA	.SDVAGHGPQ	LDAILDQL.T	Et/mal-CoA
atmon01p	SSTVISGPPE	QVAAVVADAE	ARELRGRVID	.VDYASHSPQ	VDAITDEL.T	
atdebs02p	DAVVVAGDAQ	AAREFLEYCE	GVGIRARAIP	.VDYASHTAH	VEPVRDEL.V	
atdebs06p	SSVVVSGDPE	ALAELVARCE	DEGVRAKTLP	.VDYASHSRH	VEEIRETI.L	
atave01p	RSTAVSGDAE	AVDEVLAYCA	GTGVRARRIP	.VDYASHCPH	VQPLREEL.L	
atave07p	RSTAVSGDAE	AVDEVLAYCA	GTGVRARRIP	.VDYASHCPH	VQPLREEL.L	
atave06p	HSTTVSGDTK	AVDEVLAHCT	DTGLRAKRIP	.VDYASHCPH	VQPLHDEL.L	
atave09p	HSTTVSGDTT	AVEELLTHCA	DTGLRAKRIP	.VDYASHCPH	VQPLHDEL.L	
atnys01p	RSVVVAGEPE	ALDALHARLT	ADDIRARRIA	.VDYASHSHQ	VEDLHEEL.L	
atnys11p	RSVVVAGEPE	ALDALHARLT	ADDIRARRIA	.VDYASHSHQ	VEDLHEEL.L	
atrif05p	ASVVIAGDAE	ALTEAVEVLG	G.....RRVA	.VDYASHTRH	VEDIQDTL.A	
atrif07p	ASVVIAGDAQ	ALDEALEVLA	GDGVRVRQVA	.VDYASHTRH	VEDIRDTL.A	
atrif08p	SSVVIAGDAE	ALDQALEALT	GQDIRVRRVA	.VDYASHTRH	VEDIQEPL.A	
atrif10p	ASVVIAGDAQ	ALDETLEALS	GAGIRARRVA	.VDYASHTRH	VEDIEDTL.A	
atrif03p	SSVVIAGDAQ	ALDEALEALA	GDGVRVRRVA	.VDYASHTRH	VEAIAETL.A	
atrif06p	ASVVIAGEAQ	ALDEVVDALS	GQEVRVRRVA	.VDYGSHTNQ	VEAIEDLL.A	
atrif04p	TSVVIAGDAE	ALDEALDALD	DQGVRIRRVA	.VDYASHTRH	VEAARDAL.A	
atrif01p	SSVVIAGDAH	ALDATLEILS	GEGIRVRRVA	.VDYASHTRH	VEDIRDTL.A	
atnys02p	SSVVVSGDTD	ALDALHTACQ	EQGVRARKVS	.VDYASHGRH	VEAVRDEL.A	
atfkb02p	ASIVVAGAAD	AVEELLAATP	....HARRIA	.VDYASHTAH	VESIRGAL.L	
atave11p	RSAVVSGEPE	AVDALVEELS	HEDVPARRLM	.VDWASHSPQ	VEAIQGRL.L	
atdebs03p	RSVVVSGEPG	ALRAFSEDCA	AEGIRVRDID	.VDYASHSPQ	IERVREEL.L	
atnid04p	ETTVVCGAPG	AVDSLLGVLQ	GEGVRVRRID	.VDYASHSRH	VEGIRDEL.A	
atdebs05p	RSVVVAGESG	PLDELIAECE	AEGITARRIP	.VDYASHSPQ	VESLREEL.L	
atdebs04p	GTSVVAGPTA	ELDEFFAEAE	AREMKPRRIA	.VRYASHSPE	VARIEDRL.A	

Fig 2m

```
atave02a  TSLVISGTPH TVQHITTLCQ QQGIKTKTL. PTNHAFHSPH TNPILNQLH.
atave05a  TSLVISGTPH TVQHITTLCQ QQGIKTKTL. PTNHAFHSPH TNPILNQLH.
atave04a  TSLVISGTPH TVQHITTLCQ QQGIKTKTL. PTKNAFHSPH TNPILNQLH.
atave08a  TSLVISGTPH TVQHITTLCQ QQGIKTKTL. PTNHAFHSPH TNPILNQLH.
atave03a  TSLVISGTPH TVQHITTLCQ QQGIKTKTL. PTNHAFHSPH TNPILNQLH.
atrap02a  SSVVLSGDEA AVLQAAEGLG ....KWTRL. PTSHAFHSAR MEPMLEEFR.
atrap11a  SSVVLSGDEA AVLQAAEGLG ....KWTRL. ATSHAFHSAR MEPMLEEFR.
atrap08a  SSVVLSGDEA AVLQAAEGLG ....KWTRL. ATSHAFHSAR MEPMLEEFR.
atrap12a  SSVVLSGDEA AVLQAAEGLG ....KWTRL. ATSHAFHSAR MEPMLEEFR.
atrap05a  SSVVLSGDET AVLQAAAALG ....KSTRL. ATSHAFHSAR MEPMLEEFR.
atrap09a  SSVVLSGDEA AVLQAAEGLG ....KWTRL. ATSHAFHSAR MEPMLEEFR.
atfkb03a  ASIVLSGDED AVLDVAARLG ....RFTRL. RTSHAFHSAR MEPMLDEFR.
atfkb07x  HSVVLSGDEG PVLDVAQQLG ....IHHRL. PTRHAGHSAR MDPLVAPLL. MeOmalonyl-CoA
atfkb08x  HSVVLSGDED AVLDVAQRLG ....IHHRL. PAPHAGHSAH MEPVAAELL. MeOmalonyl-CoA
atnid01a  THCVLSGPRT ALEETAQQLH QQGIRHTWL. KVSHAFHSAL MDPMLGAFR.
atnid03a  THCVLSGPRT ALEETAQHLR EQNVRHTWL. KVSHAFHSAL MDPMLGAFR.
atnid02a  THCVLSGPRT ALEETAQHLR EQNVRHTWL. KVSHAFHSAL MDPMLGAFR.
atnid00a  THCVLSGPRT ALEETAQHLR EQNVRHTWL. KVSHAFHSAL MDPMLGAFR.
atfkb10a  SAVVLTGAPD DVAAFEREWA AAGRRAKRL. DVGHAFHSRH VDGALDDFR.
atrap14a  EAVVVSGEPE PVADFEAAWT ASGREARKL. KVRHAFHSRH VEAVLDEFR.
atmon06a  DSTVISGPSD EVDRIAGVWR ERGRKTKAL. SVSHAFHSAL MEPMLAEFT.
atmon08a  DSTVISGPSG EVDRIAGVWR ERGRKTKAL. SVSHAFHSAL MEPMLAEFT.
atmon09a  DSTVISGPSG EVDRIAGVWR ERGRKTKAL. SVSHAFHSAL MEPMLGEFT.
atepo02a  EQVVIAGVEQ AVQAIAAGFA ARGARTKRL. HVSHAFHSPL MEPMLEEFG.
atepo03x  EQVVIAGVEQ AVQAIAAGFA ARGARTKRL. HVSHASHSPL MEPMLEEFG. Mal/mmal
atepo08a  EQVVIAGAEK FVQQIAAAFA ARGARTKPL. HVSHAFHSPL MDPMLEAFR.
atepo00a  DQVVIAGAGQ PVHAIAAAMA ARGARTKAL. HVSHAFHSPL MAPMLEAFG.
atepo04a  DAVVIAGAEV QVLALGATFA ARGIRTKRL. AVSHAFHSPL MDPMLEDFQ.
atnid07a  DSLVLSGDEQ AVVSAAGELA ARGRRTKRL. SVSHAFHSPH MDAMLADFR.
atty107a  RSVVISGAEE AVAEAAAQLA GRGRRTRRL. RVAHAFHSPL MDGMLAGFR.
atsor02a  LSTVVAGDED AVVEIARQAE ALGRKTTRL. RVSHAFHSPH MDGMLDDFR.
atsorb1a  LSTVVAGDED AVLKIARQVE ALGRKATRL. RVSHAFHSPH MDGMLDDFR.
atnys09a  SAVVLSGAEA TVTALAEQLA ADGRKTRRL. RVSHAFHSPL MEPMLDAFR.
atnys12a  RSVVVAGVEE DVLLLADLFA ADGRRTKRL. RVSHAFHSPL MDAMLDDFA.
atnys16a  VSVVVSGVEA AVGQVVDQLV ERGRRVRRL. AVSHAFHSPL MDPMLDAFR.
atnys17a  VSVVVSGVEA AVGQVVDQLV ERGRRVRRL. AVSHAFHSPL MDPMLDAFR.
atnys03a  TSVVVAGTEE AVAAIGARFT AQDRKTTRL. RVSHAFHSPL MDPMLAEFR.
atnys15a  TSLVVSGDET ATLAVAARLA EQGRRTTRL. RVSHAFHSPL MDPMLAEFR.
atnys07a  NALVVSGVED AAVEIGARFA AEGRRTTRL. HVSHAFHSPL MDPMLAEFR.
atnys08a  TSVVISGAEE ATQTVAQHFA DQGRRTTAL. RVSHAFHSPL MDPMLAEFR.
atnys05a  TSVVVSGAES AARTVADRLA ENGRKTTRL. RVSHAFHSPL MDPMLAEFR.
atnys06a  TSVVISGAEE ATQTVAQHFA DQGRRTTAL. RVSHAFHSPL M..MLAEFR.
atnys04a  TSVVVSGYEN ATLAVARHFA DQGRRTTRL. RVSHAFHSPL MAPMLDDFR.
atnys14a  TAVVLSGAGD AVTALGQALA ERGHRTTRL. RVSHAFHSHL MDPMLADFR.
atnys00a  TAVVVAGAED AVRQLTARFA DRGRRTSRL. AVSHAFHSPL MEPMLDAFR.
atnys10a  QSVVISGDEE AAETIAATFA ERGRKTKRL. RVSHAFHSPR MDGMLDAFR.
atnys18a  RSVVVAGAED AALAVRRHFD DLGRRTTRL. PVSHAFHSPL MDPMLDAFR.
atnys13a  RSVVVAGAED AVRAVADRLA ADGRRTRRL. TVSHAFHSPL MDPMLTDFA.
atave10a  RSIVLSGDED AVLDLAQQWA ARGRRTRRL. RTSHAFHSPH MDAMLGDFR.
atrif02a  DSVVLSGTEA AVLAVADELA GRGRKTRRL. AVSHAFHSPL MEPMLDDFR.
atmon03a  TAVVVSGEAA AVGEVEKALR GRGLKTKRL. NVSHAFHSPL IEPMLDDFR.
atave12a  ASVVFSGAED EVGNMADWFA ERGRRVKRL. RTGHAFHSPL MDPMLEEFQ.
atrif09a  SAVVLSGDAD AVVAAAARMR ERGHKTKQL. KVSHAFHSAR MAPMLAEFA.
atmon00a  DSVVVSGDRA TVDELTAAWR GRGRKAHHL. KVSHAFHSPH MDPILDELR.
atty103a  SSVVVSGDEA AVLELLEQWR AEGREARRL. AVSHAFHSPR MDGMLTQFD.
                                          ****  HAFH/YASH/TAGH motif
```

Fig 2n

```
          351                                                      400
 atave00x SGLLPITPRP SRIPFHSSVT G.....GRL. .DTRELDAAY WYRNMSSTVR
atdebs00p SALAWFAPGG SEVPFYASLT G.....GAV. .DTRELVADY WRRSFRLPVR
 atepo06p AALGAIRPRA AAVPMRSTVT G.....GVI. .AGPELGASY WADNLRQPVR
 atepo07p AALGAIRPRA AAVPMRSTVT G.....GVI. .AGPELGASY WADNLRQPVR
 atepo01p AALGGLRPGA AAVPMRSTVT G.....AMV. .AGPELGANY WMNNLRQPVR
 atepo05p AALGELEPRQ ATVSMRSTVT S.....TIM. .AGPELVASY WADNVRQPVR
 atsora1x DALREVRPNK AQIPIVSEVT G.....TAL. .DGERFDASH WVRNFGDPAL
 atfkb01p DALEGITSST PSVPWWSTVD S.....GWV. ..TEPFGDAY WYRNLRQPVA
 atfkb09p DVLGDITSSA PSVPWWSTVD G.....GWV. ..TEPAGDDY WYRNLRQPVA
 atrap03p DITSDSSSQA PLVPWLSTVD G.....SWV. ..DSPLDGEY WYRNLREPVG
 atrap06p DITSDSSSQA PVVPWLSTVD G.....SWV. ..DSPLDVEY WYRNLREPVG
 atrap04p GITAGIGSQP PVVPWLSTVD G.....SWV. ..DSPLDGEY WYRNLREPVG
 atrap13p DITSDSSSQT PLVPWLSTVD G.....TWV. ..DSPLDGEY WYRNLREPVG
 atrap01p GVIAGVDSRA PVVPWLSTVD G.....TWV. ..EGPLDAEY WYRNLREPVG
 atrap07p DITAGIGSQA PVVPWLSTVD G.....TWV. ..EGPLDVEY WYRNLREPVG
 atrap10p DITSDSSSQD PLVPWLSTVD G.....TWV. ..DSPLDGEY WYRNLREPVG
 atfkb04x RIVAATTSRA PEIPWFSTAD E.....RWI. ..DAPLDDEY WFRNMRNPVG
 atty104p RVLSGIRPRS PRVPVCSTVA G.....E..Q PGEPVFDAGY WFRNLRNRVE
 atty106p RVLSGIRPRS PRVPVCSTVA G.....E..Q PGEPVFDAGY WFRNLRNRVE
 atty101p RVLSGIRPRS PRVPVCSTVA G.....E..Q PGEPVFDAGY WFRNLRNRVE
 atty102p RVLSGIRPRS PRVPVCSTVA G.....E..Q PGEPVFDAGY WFRNLRNRVE
 atty100p RVLSGIRPRS PRVPVCSTVA G.....E..Q PGEPVFDAGY WFRNLRNRVE
 atnid05b EVLAPVAPRP ADIPFYSTVT G.....GLL. .DGTELDATY WYRNMREPVE
 atty105b EVLAPVSPRS SDIPFYSTVT G.....APL. .DTERLDAGY WYRNMREPVE
 atnid06x DLLSGVRPAP SRIPFFSTVT A.....GPC. .GGDQLDGAY WYRNTREPVE
atdebs01p ELGEDFHPLP GFVPFFSTVT G.....RWT. .QPDELDAGY WYRNLRRTVR
 atmon02p ERLADIRPTN TDVAFYSTVT A.....ERL. TDTTALDTDY WVTNLRQPVR
 atmon10p ERLADIRPAN TDVAFYSTVT A.....ERL. TDTTALDTDY WVTNLRQPVR
 atmon04p EGLADIRPAN TDVAFYSTVT A.....ERL. TDTTALDTDY WVTNLRQPVR
 atmon07p DRLADIRPAT TDVAFYSTVT A.....ERL. TDTTALDTDY WVTNLRQPVR
 atmon11p DRLADIQPTT TDVAFYSTVT A.....ERL. DDTTALDTAY WVTNLRQPVR
 atmon12p ERLADIRPTT TDVAFYSTVT A.....ERL. DDTTTLDTDY WVTNLRQPVR
 atmon05b EGLAGIRPAA TDVAFYSTVT A.....GHL. TDTTELDTAY WVRNVRRTVR
 atmon01p HTLSGVRPTT APVAFYSAVT G.....TRI. .DTAGLDTDY WVTNLRRPVR
atdebs02p QALAGITPRR AEVPFFSTLT G.....D..F LDGTELDAGY WYRNLRHPVE
atdebs06p ADLDGISARR AAIPLYSTLH G.....E..R RD...MGPRY WYDNLRSQVR
 atave01p ELLGDISPQP SGVPFFSTVE G.......TW LDTTTLDAAY WYRNLHQPVR
 atave07p ELLGDISPQP SGVPFFSTVE G.......TW LDTTTLDAAY WYRNLHQPVR
 atave06p HLLGDITPQP STVPFFSTVE G.......TW LDTTTLDAAY WYRNLHQPVR
 atave09p HLLGDITPQP STMPFFSTVV G.....HLVW Y.TTTLDAAY WYRNLHQPVR
 atnys01p EVLAELAPRT SEVPFFSTVT G.....DWL. .DTARMDAGY WFRNLRGRVR
 atnys11p EVLAELAPRT SEVPFFSTVT G.....DWL. .DTARMDAGY WFRNLRGRVR
 atrif05p ETLAGIDAQA PVVPFYSTVA G.....EWI. TDAGVVDGGY WYRNLRNQVG
 atrif07p ETLAGITAQA PDVPFRSTVT G.....GWV. RDADVLDGGY WYRNLRNQVR
 atrif08p EALAGIEAHA PTLPFFSTLT G.....DWI. REAGVVDGGY WYRNLRNQVG
 atrif10p EALAGIDARA PLVPFLSTLT G.....EWI. RDEGVVDGGY WYRNLRGRVR
 atrif03p KTLAGIDARV PAIPFYSTVL G.....TWI. EQA.VVDAGY WYRNLRQQVR
 atrif06p ETLAGIEAQA PKVPFYSTLI G.....DWI. RDAGIVDGGY WYRNLRNQVG
 atrif04p EMLGGIRAQA PEVPFYSTVT G.....GWV. EDAGVLDGGY WYRNLRRQVR
 atrif01p ETLAGISAQA PAVPFYSTVT S.....EWV. RDAGVLDGGY WYRNLRNQVR
 atnys02p RVLAPVDPRA PEVPFYSTVT G.....DRV. DDAA.FDGAY WYTNLRQTVR
 atfkb02p DALADLTPGA PEIPFFSTVD E.....AWL. DRPA..DAAY WYDNVRCPVR
 atave11p ELLAPIRART GDVPFYSTVT G.....ERI. .DGTELDADY WYRNLRQVVR
atdebs03p ETTGDIAPRP ARVTFHSTVE S.....RSM. .DGTELDARY WYRNLRETVR
 atnid04p AVLAGLRPRA GRVPFYSTVE A.....EPL. .DGTALDAGY WYRNLRQRVR
atdebs05p TELAGISPVS ADVALYSTTT G.....QPI. .DTATMDTAY WYANLREQVR
atdebs04p AELGTITAVR GSVPLHSTVT G.....EVI. .DTSAMDASY WYRNLRRPVL
```

Fig 2o

```
atave02a  QHTQTLTYHP PHTPLITANT .......... PPDQLLTPHY WTQQARNTVD
atave05a  QHTQTLTYHP PHTPLITANT .......... PPDQLLTPHY WTQQARNTVD
atave04a  QHTQTLTYHP PHTPLITANT .......... PPDQLLTPHY WTQQARNTVD
atave08a  QHTQTLTYHP PHTPLITANT .......... PPDQLLTPHY WTQQARNTVD
atave03a  QHTQTLTYHP PHTPLITANT .......... PPDQLLTPHY WTQQARNTVD
atrap02a  AVAEGLTYRT PQVA..... .........MA AGDQVMTAEY WVRQVRDTVR
atrap11a  AVAEGLTYRT PQVS..... .........MA VGDQVTTAEY WVRQVRDTVR
atrap08a  AVAEGLTYRT PQVS..... .........MA AGDQLTTTEY WVRQVRDTVR
atrap12a  AVAEGLTYRT PQVS..... .........MA VGDQVTTAEY WVRQVRDTVR
atrap05a  TVAERLTYQT PRLA..... .........MA AGDRVTTAEY WVRQVRDTVR
atrap09a  AVAQGLTYHA PGVV..... .........MA AGDRVMTAEY WVRQVRDTVR
atfkb03a  DVAERLTYHE PKLP..... .........MA AGADCATPEY WVRQVRDTVR
atfkb07x  EAASGLTYHQ PHT...... ..........A IPEDPTTAAY WARQVRDQVR
atfkb08x  ATTRELRYDR PHT...... ..........A IPNDPTTAEY WAEQVRNPVL
atnid01a  DTLNTLNYQP PTIPLISNLT GQIADPNHL. .....CTPDY WIDHARHTVR
atnid03a  DTLNTLNYQP PTIPLISNLT GQIADPNHL. .....CTPDY WIDHARHTVR
atnid02a  DTLNTLNYQP PTIPLISNLT GQIADPNHL. .....CTPDY WIDHARHTVR
atnid00a  DTLNTLNYQP PTIPLISNLT GQIADPNHL. .....CTPDY WIDHARHTVR
atfkb10a  GVLESLAFGA ARLPVVSTTT GRDAAGD.LA ......TPEH WLRHARRPVL
atrap14a  TALESLKFRA PALPVVSTVT GRLIDQDEMG ......TPEY WLRQVRRPVR
atmon06a  EAIRGVKFRQ PSIPLMSNVS GERA...... .GEEITDPEY WARHVRNAVL
atmon08a  EAIREVKFTR PKVSLISNVS GLEA...... .GEEIASPEY WARHVRQTVL
atmon09a  EAIRGVKFRQ PSIPLMSNVS GERA...... .GEEITSPEY WARHVRQTVL
atepo02a  RVAASVTYRR PSVSLVSNLS GKVVT.DEL. .....SAPGY WVRHVREAVR
atepo03x  RVAASVTYRR PSVSLVSNLS GKVVA.DEL. .....SAPGY WVRHVREAVR
atepo08a  RVTESVTYRR PSIALVSNLS GKPCT.DEV. .....SAPGY WVRHAREAVR
atepo00a  RVAESVSYRR PSIVLVSNLS GKACT.DEV. .....SSPGY WVRHAREVVR
atepo04a  RVAATIAYRA PDRPVVSNVT GHVAG.PEI. .....ATPEY WVRHVRSAVR
atnid07a  AVAESVTYRT PRLPIVSEVT GRPAAPSEL. .....MDPGY WTRQIREPVR
atty107a  EVAAGLRYRE PELTVVSTVT GRPARPGEL. .....TGPDY WVAQVREPVR
atsor02a  RVAQSLTYHP ARIPIISNVT GARATDHEL. .....ASPDY WVRHVRHTVR
atsorb1a  RVAQGLTFHP ARIPIISNVT GARATDQEL. .....ASPET WVRHVRDTVR
atnys09a  AVVEDLTLQP PLLPVVSNLT GKPATVAQL. .....TSADY WVDHVRHAVR
atnys12a  AVARGLTYHP PTIPFVSNVS GGLATAEQV. .....RTPDY WVGHVRAAVR
atnys16a  AVAEGLEYHQ PRIPVVSNVT GEVAAAEEL. .....CAADY WVRHVRATVR
atnys17a  AVAEGLEYHQ PRIPVVSNVT GEVAAAEEL. .....CAADY WVRHVRATVR
atnys03a  AVAAGLTYHE PRIPVLSNLT GTVAAVADL. .....CSADY WVRHVREAVR
atnys15a  AVAEGLSYGE PQIPVVSNLT GAVADGTLL. .....GTADY WVRHVREAVR
atnys07a  VVAEGLSYAA PSLPVVSNLT GQVATADEL. .....CSAEY WVRHVREAVR
atnys08a  AVAEGLSYAT PSLPVVSNLT GWLATADEL. .....CSAEY WVRHVREAVR
atnys05a  AVAEGLSYAT PTLPVVSNLT GRLATADDL. .....CSAEY WARHVREAVR
atnys06a  AVAEGLSYAT PTLPVVSNLT GQVATADEL. .....CSAEY WVRHVREAVR
atnys04a  AVVESLTFTA PTTPVVSNLT GELAPAEAL. .....CSADY WVRHVREAVR
atnys14a  TVAEGLEYHP PRIPVVSNLT GDVADAADL. .....CSADY WVRHVRGTVR
atnys00a  DVVSRLTFHQ PSIPLVSNLT GELA.GSEI. .....TSAEY WVRHVRDTVR
atnys10a  IVAEGLTYRA PRIPLVSDLT GRRADDAEV. .....CTAEY WVRHVREAVR
atnys18a  TALAPLTFAE PEIPVVSNLT GLPATAEEL. .....ATPHY WVCHVRQAVR
atnys13a  RVAEGLTYHE PRIPLVSTLL GAPAGA.EL. .....RTPDY WVRHVRETVR
atave10a  RAAEQVTFSA PRIPVVSNVT GAPLPAETM. .....CTPDY WVEHARSTVR
atrif02a  AVAERLTYRA GSLPVVSTLT GELAA...L. .....DSPDY WVGQVRNAVR
atmon03a  EVARGLTFHA PTLPVVSNLT GRLADAELM. .....ADAEY WVRHVRRPVR
atave12a  QVAASLTYSE PAIPMVSTLT GDIVAAGEL. .....SDPEY WVRQVRRTVR
atrif09a  AELAGVTWRE PEIPVVSNVT GRFAEPGEL. .....TEPGY WAEHVRRPVR
atmon00a  AVAAGLTFHE PVIPVVSNVT GELVTATATG SGAGQADPEY WARHAREPVR
atty103a  RVARTLTFAP PTIPLVSTLT GTPVTEETL. .....CTADH WVRQAREPVR
```

Fig 2p

```
          401                                                  450
 atave00x  FEPAARLLLQ QGP.KTFVEM SPHPVLTMGL QELAPDLG.. ......DTTG
atdebs00p  FDEAIRSALE VGP.GTFVEA SPHPVLAAAL QQTL...... ......DAEG
 atepo06p  FAAAAQALLE GGP.ALFIEM SPHPILVPPL DEIQTA.... ........AE
 atepo07p  FAAAAQALLE GGP.ALFIEM SPHPILVPPL DEIQTA.... ........AE
 atepo01p  FAEVVQAQLQ GGH.GLFVEM SPHPILTTSV EEMRRA.... ........AQ
 atepo05p  FAEAVQSLME DGH.GLFVEM SPHPILTTSV EEIRRA.... ........TK
 atsora1x  FSTAIDHLLQ EGF.DIFLEL TPHPLALPAI ESNLRR.... ........SG
 atfkb01p  MDTAVSELDG ....SLFIEC SAHPVLLPAL DQ........ ..........
 atfkb09p  MDTAIGELDG ....SLFIEC SAHPVLLPAL DQ........ ..........
 atrap03p  FHPAVGQLQA QGD.TVFVEV SASPVLLQAM DD........ ..........
 atrap06p  FHPAVGQLQA EGD.TVFVEV SASPVLLQAM DD........ ..........
 atrap04p  FHPAVSQLQA QGD.AVFVEV SASPVLLQAM DD........ ..........
 atrap13p  FHPAVSQLQA QGD.TVFVEV SASPVLLQAM DD........ ..........
 atrap01p  FEPAAGQLQA QGD.TVFVEV SASPVLLQAM DD........ ..........
 atrap07p  FDSAVGQLRA EGD.TVFVEV SASPVLLQAM DD........ ..........
 atrap10p  FHPAVSQLQA QGD.TVFVEV SASPVLMQAM DD........ ..........
 atfkb04x  FAAAVAAARE PGD.TVFIEV SAHPVLLPAI NG........ ..........
 atty104p  FSAVVGGLLE EGH.RRFIEV SAHPVLVHAI EQT....A.. ....EAAD
 atty106p  FSAVVGGLLE EGH.RRFIEV SAHPVLVHAI EQT....A.. ......EAAD
 atty101p  FSAVVGGLLE EGH.RRFIEV SAHPVLVHAI EQT....A.. ......EAAD
 atty102p  FSAVVGGLLE QGH.RRFIEV SAHPVLVHAI EQT....A.. ......EAAD
 atty100p  FSAVVGGLLE EGH.RRFIEV SAHPVLVHAI EQT....A.. ......EAAD
 atnid05b  FERATRALIA DGH.DVFLET SPHPMLAVAL EQT....V.. ......TDAG
 atty105b  FEKAVRALIA DGY.DLFLEC NPHPMLAMSL DET....L.. ......TDSG
 atnid06x  FDATVRALLR AGH.HTFIEV GPHPLLNAAI DEI....A.. ......ADEG
atdebs01p  FADAVRALAE QGY.RTFLEV SAHPILTAAI EEI....G.. ......DGSG
 atmon02p  FADTIEALLA DGY.RLFIEA SAHPVLGLGM EETIEQ.... ........AD
 atmon10p  FADTIEALLA DGY.RLFIEA SAHPVLGLGM EETIEQ.... ........AD
 atmon04p  FADTIEALLA DGY.RLFIEA SAHPVLGLGM EETIEQ.... ........AD
 atmon07p  FADTIDALLA DGY.RLFIEA SAHPVLGLGM EETIEQ.... ........AD
 atmon11p  FADTIEALLA DGY.RLFIEA SPHPVLNLGI QETIEQQA.. .......GAA
 atmon12p  FADTIEALLA DGY.RLFIEA SPHPVLNLGM EETIER.... ........AD
 atmon05b  FADTIDALLA DGY.RLFIEV SPHPVLNLAL EGLIER.... ........AA
 atmon01p  FADAVTALLA DGH.RVFIEA SSHPVLTLGL QETFEE.... ........AG
atdebs02p  FHSAVQALTD QGY.ATFIEV SPHPVLASSV QETL...... ......DDAE
atdebs06p  FDEAVSAQSP DGH.ATFVEM SPHPVLTAAV QE........ ........IA
 atave01p  FSDAVQALAD DGH.RVFVEV SPHPTLVPAI EDTTEDTA.. ......ED..
 atave07p  FSDAVQALAD DGH.RVFVEV SPHPTLVPAI EDTTEDTA.. ......ED..
 atave06p  FSHAIQTLTD DGH.RAFIEI SPHPTLVPAI EDTTENTT.. ......EN..
 atave09p  FSHAIQTLTD DGH.RPFIEI SPHPTLVPAI EDTTENTT.. ......EN..
 atnys01p  FADAVADLLA AEY.RAFVEV SSHPVLTMAV LD....LI.. ......EEAG
 atnys11p  FADAVADLLA AEY.RAFVEV SSHPVLSMAV QE....AI.. ......DEAG
 atrif05p  FGPAVAELIE QGH.GVFVEV SAHPVLVQPI SE....LT.. ......D...
 atrif07p  FGPAVAELLE QGH.GVFVEV SAHPVLVQPI SE....LT.. ......D...
 atrif08p  FGPAVAELLG LGH.RVFVEV SAHPVLVQAI SA....IA.. ......DD..
 atrif10p  FGPAVEALLA QGH.GVFVEL SAHPVLVQPI TE....LT.. ......DE..
 atrif03p  FGPSVADLAG LGH.TVFVEI SAHPVLVQPL SE....IS.. ......DD..
 atrif06p  FGPAVAELVR QGH.GVFVEV SAHPVLVQPL SE....LS.. ......DD..
 atrif04p  FGPAVAELIE QGH.RVFVEV SAHPVLVQPI NE....LV.. ......DD..
 atrif01p  FGAAATALLE QGH.TVFVEV SAHPVTVQPL SE....LT.. ......GD..
 atnys02p  MEEATRALLA AGH.RVFIEV SPHPVLAAPI QETQEAVA.. ......EATG
 atfkb02p  FGAAAARLAE LGH.RVFVEA SPHPVLTTAL ADTLAG.... .........H
 atave11p  FRDATQALVR AGH.TVFIEA CPHPAVAVGV QETLDE.M.. ........GD
atdebs03p  FADAVTRLAE SGY.DAFIEV SPHPVVVQAV EEAVEE.A.. ......DGAE
 atnid04p  FESALRAMLA DGV.DAFVEC SPHPVLTVPV RQTLED.A.. .......GA.
atdebs05p  FQDATRQLAE AGF.DAFVEV SPHPVLTVGI EATLDS.A.. ......LPAD
atdebs04p  FEQAVRGLVE QGF.DTFVEV SPHPVLLMAV EET....A.. ......EHAG
```

Fig 2q

```
atave02a  YATTTQTLHQ HG.VTTYIEL GPDNTLTTLT HHNLPNPPTT TLTLTHPHHH
atave05a  YATTTQTLHQ HG.VTTYIEL GPDNTLTTLT HHNLPNTPTT TLTLTHPHHH
atave04a  YATTTQTLHQ HG.VTTYIEL GPDNTLTTLT HHNLPNTPTT TLTLTHPHHH
atave08a  IATTTQTLHQ HG.VTTYIEL GPDNTLTTLT HHNLPNTPTT TLTLTHPHHH
atave03a  YATTTQTLHQ HG.VTTYIEL GPDNTLTTLT HDNLPNTPTT TLTLTHPHHH
atrap02a  FGEQVASFED A....VFVEL GADRSLARLV DG..... .....
atrap11a  FGEQVASYED A....VFVEL GADRSLARLV DG..... .....
atrap08a  FGEQVASYED A....VFVEL GADRSLARLV DG..... .....
atrap12a  FGEQVASYED A....VFVEL GADRSLARLV DG..... .....
atrap05a  FGEQVASYED A....VFIEL GADRSLARLV DG..... .....
atrap09a  FGEQVASYED A....VFVEL GADRSLARLV DG..... .....
atfkb03a  FAEQVAAYDG A....ALLEI GPDRNLARLV DG..... .....
atfkb07x  FQAHAERYPG A....TFLEI GPNQDLSPVV DG..... .....
atfkb08x  FHAHTQRYPD A....VFVEI GPGQDLSPLV DG..... .....
atnid01a  FADAVQTAHD QR.TTTYLEI GAHPTLTTLL HHTLDNP... .....
atnid03a  FADAVQTAHH QG.TTTYLEI GPHPTLTTLL HHTLDNP... .....
atnid02a  FADAVQTAHD QR.TTTYLEI GPHPTLTTLL HHTLDNP... .....
atnid00a  FADAVQTAHH QG.TTTYLEI GPHPTLTTLL HHTLDNP... .....
atfkb10a  YADAVRELAD LG.VNMFVAV GPSGALASAA SENTGGSAGT YH.....
atrap14a  FQDAVRELAE QG.VGTFVEV GPSGALASAG VECLGGDA.S FH.....
atmon06a  FQPAIAQVAD S..AGVFVEL GPAPVLTTAA QHTLDE.SD. .SQES.....
atmon08a  FQPGIAQVAS T..AGVFVEL GPGPVLTTAA QHTLDDVTDR HGPEP.....
atmon09a  FQPGVAQVAA E..ARAFVEL GPGPVLTAAA QHTLDHITEP EGPEP.....
atepo02a  FADGVKALHE AG.AGTFVEV GPKPTLLGLL PACLPEAEP. .....
atepo03x  FADGVKALHE AG.AGTFVEV GPKPTLLGLL PACLPEAEP. .....
atepo08a  FADGVKALHA AG.AGLFVEV GPKPTLLGLV PACLPDARP. .....
atepo00a  FADGVKALHA AG.AGTFVEV GPKSTLLGLV PACMPDARP. .....
atepo04a  FGDGAKALHA AG.AATFVEV GPKPVLLGLL PACLGEADA. .....
atnid07a  FAAAVRAARA AG.AATFVEL GPDAVLSGMA RECAAG.... .....DTGT
atty107a  FADAVRTAHR LG.ARTFLET GPDGVLCGMA EECLED.... .....DTVA
atsor02a  FLDGVRALHA EG.ARVFLEL GPHAVLSALA QDALGQ.... ....D.EGTS
atsorb1a  FLDGVRTLHA EG.ARAFLEL GPHPVLSALA QDALGH.... ....D.EGPS
atnys09a  FADGIDWLA. RHDTTAFLEL GPDGVLSAMA QDCLDA.... ....A.DAD.
atnys12a  FADGIDWLAT QGDVHTFLEL GPDGVLSAMA RESLTD.... ....P.SRT.
atnys16a  FADGVRTLAE RG.ATAFLEI GPDGVLSALA RGVL..... ....P.AEA.
atnys17a  FADGVRTLAE RG.ATAFLEI GPDGVLSALA AACL.F.... ....D.TDA.
atnys03a  FADGVTALTD RG.VTTLVEL GPDGVLSAMA QESL..... ....P.DGA.
atnys15a  FADGIRALTD AG.VGAFLEL GPDGTLAALA QQSA..... ....P.D.A.
atnys07a  FADGVTALEA EG.VRTFLEL GPDGVLAAMA GASL..... ....T.ESS.
atnys08a  FADGITTLEA EG.VRTFLEL GPDGILSALA QQSL..... ....A.GEA.
atnys05a  FADGVSTLEN EG.VTTFLEL GPDGVLSAMA QQSL..... ....T.GDA.
atnys06a  FADGVTALEA EG.VRTFLEL GPDGVLAAMA RETV..... ....A.DDT.
atnys04a  FADGIRTLAD RG.VTTFVEL GPDSVLSAMA QESA..... ....P.EGA.
atnys14a  FADGVRTMAD RG.VHLFLEL GPDAVLSAMA RQCA..... ....P.D.A.
atnys00a  FADGITALAK AG.ADVLIEL GPGGVLSAMA RDTL.G.... ....P.DST.
atnys10a  FADCVRTLRD AG.ATTFLEL GSDGLLTAMA EDTL.G.... ....D.DHD.
atnys18a  FGDGVRALAD RG.VRTFLEL GPDGVLSALV RENL..... ....P.EPG.
atnys13a  FADGVRALHD AG.AGTFVEI GPDGVLTALT QQTLDT.... ....V.EAGA
atave10a  FADGISWLQE QG.VTTCLEI GPDGTLSALA QDSLSA.... ....P.....
atrif02a  FSDAVTALGA QG.ASTFLEL GPGGALAAMA LGTLGG.... ....P.EQSC
atmon03a  FHDGLRALSE QGVVR.YLEL GPDPVLATMV QDGLPA.... ....P.AEGE
atave12a  FGDAISRLHT DG.VRTFMEL GPDGTLSALA EECLEATADS HPADD.DTGT
atrif09a  FAEGVAAATE SGG.SLFVEL GPGAALTALV EET..... .....
atmon00a  FLSGVRGLCE RG.VTTFVEL GPDAPLSAMA RDCFPAPADR SRPRP.....
atty103a  FLDAMRTLRA DG.IDTFVEL GPDGVLSAMA RDCADDRPDG DTTGAGDGET
```

Fig 2r

```
                451                                                  500
 atave00x   TADTVIMGTL RRGQGTLDHF LTSLAQLRGH GE..TSATTV LSARLTALSP
atdebs00p   SSAAVV.PTL QRGQGGMRRF LLAAAQAFTG GV..AVDWTA AYDDVGA.EP
 atepo06p   QGGAAV.GSL RRGQDERATL LEALGTLWAS G..YPVSWAR LFPAGG....
 atepo07p   QGGAAV.GSL RRGQDERATL LEALGTLWAS G..YPVSWAR LFPAGG....
 atepo01p   RAGAAV.GSL RRGQDERPAM LEALGTLWAQ G..YPVPWGR LFPAGG....
 atepo05p   REGVAV.GSL RRGQDERLSM LEALGALWVH G..QAVGWER LFSAGGAGL.
 atsora1x   RRGVVL.PSL RRNEDERGVM LDTLGVLYVR G..APVRWDN VYPA...AF.
 atfkb01p   .E.RTV.ASL RTDDGGWDRF LAALAQAWTQ GA..DVDWTT LIEPA.....
 atfkb09p   .E.RTV.ASL RTDDGGWDRF LTALAQAWTQ GA..DVDWTT LIAPA.....
 atrap03p   .DVVTV.ATL RRDDGDATRM LTALAQAYVH GV..TVDWPA ILG.T.....
 atrap06p   .DVVTV.ATL RRDDGDATRM LTALAQAYVH GV..TVDWPA ILG.T.....
 atrap04p   .DVVTV.ATL RRDDGDATRM LTALAQAYVH GV..TVDWPA ILG.T.....
 atrap13p   .DVVTV.ATL RRDDGDATRM LTALAQAYVH GV..TVDWPA ILG.T.....
 atrap01p   .DVVTV.ATL RRDDGDATRM LTALAQAYVE GV..TVDWPA VLG.T.....
 atrap07p   .DVVTV.ATL RRDDGDATRM LTALAQAFVE GV..TVDWPA ILG.T.....
 atrap10p   .DVVTV.ATL RRDDGDATRM LTALAQAYVH GV..TVDWRA VLGDV.....
 atfkb04x   ...TTV.GTL RR.GGGADRV LDSLAKAHTV GV..AVDWST VVAATGAADD
 atty104p   RSVHAT.GTL RRQDDSPHRL LTSTAEAWAH G..ATLTW.. ..........
 atty106p   RSVHAT.GTL RRQDDSPHRL LTSTAEAWAH G..ATLTW.. ..........
 atty101p   RSVHAT.GTL RRQDDSPHRL LTSTAEAWAH G..ATLTW.. ..........
 atty102p   RSVHAT.GTL RRQDDSPHRL LTSTAEAWAH G..ATLTW.. ..........
 atty100p   RSVHAT.GTL RRQDDSPHRL LTSTAEAWAH G..ATLTW.. ..........
 atnid05b   TDAAVL.GTL RRRHGGPRAL ALAVCRAFAH GVE..VDPEA VF........
 atty105b   GHGTVM.HTL RRQKGSAKDF GMALCLAYVN GLE..IDGEA LF........
 atnid06x   VAATAL.HTL QRGAGGLDRV RNAVGAAFAH GVR..VDWNA LF........
atdebs01p   ADLSAI.HSL RRGDGSLADF GEALSRAFAA GVA..VDWES VH........
 atmon02p   MPATVV.PTL RRDHGDTTQL TRAAAHAFTA G..ADVDWRR WF........
 atmon10p   IPATVV.PTL RRDHGDTTQL TRAAAHAFTA G..APVDWRR WF........
 atmon04p   IPATVV.PTL RRDHGDTTQL TRAAAHAFTA G..ADVDWRR WF........
 atmon07p   IPATVV.PTL RRDHGDTTQL TRAAAHAFTA G..ATVDWRR WF........
 atmon11p   GTAVTI.PTL RRDHGDTTQL TRAAAHAFTA G..APVDWRR WF........
 atmon12p   MPATVV.PTL RRDHGDAAQL TRAAAQAFGA G..AEVDWTG WF........
 atmon05b   VPATVV.PTL RRDHGDTTQL ARAAAHAFAA G..ADVDWRR WF........
 atmon01p   VDAVTV.PTL RREDGGRARL ARSLAQAFGA G..CAVRWEN WF........
atdebs02p   SDAAVL.GTL ERDAGDADRF LTALADAHTR GVA..VDWEA VL........
atdebs06p   ADAVAI.GSL HRDTAE.EHL IAELARAHVH GVA..VDWRN VF........
 atave01p   ..VTAI.GSL RRGDNDTRRF LTALAHTHTT GIGTPTTWHH HY........
 atave07p   ..VTAI.GSL RRGDNDTRRF LTALAHTHTT GIGTPTTWHH HY........
 atave06p   ..ITAT.GSL RRGDNDTHRF LTALAHTHTT GIGTPTTWHH HY........
 atave09p   ..ITAT.GSL RRGDNDTHRF LTALAHTHTT GIRTPTTWHH HY........
 atnys01p   VTAVAT.GTL RRDQGGAGRF LLSAAEVFVR GV..DVDWAG AF........
 atnys11p   VPAVAA.GTL RRDQGGTDRF LLSAAEVFVR GV..DVDWAG LF........
 atrif05p   ..AVVT.GTL RRDDGGVRRL LTSMAELFVR GV..PVDWAT MA........
 atrif07p   ..AVVT.GTL RRDDGGLRRL LTSMAELFVR GV..RVDWAT LV........
 atrif08p   TDAVVT.GSL RREEGGLRRL LTSMAELFVR GV..DVDWAT MV........
 atrif10p   TAAVVT.GSL RRDDGGLRRL LTSMAELFVR GV..EVDWTS LV........
 atrif03p   ..AVVT.GSL RRDDGGLRRL LASAAELYVR GV..AVDWTA AV........
 atrif06p   ..AVVT.GSL RREDGGLRRL LTSMAELYVQ GV..PLDWTA VL........
 atrif04p   TEAVVT.GTL RREDGGLRRL LASAAELFVR GV..TVDWSG VL........
 atrif01p   ....AI.GTL RREDGGLRRL LASMGELFVR GI..DVDWTA MV........
 atnys02p   GSAVVL.GSL RRDEGGPRRF LTSLAEAHTH GA..PVDWTT TF........
 atfkb02p   PNTAVT.GTL RRGDGGARRF TRSLAELWVR GV..PVSW.. ..........
 atave11p   LDSLVV.GSL RRGEGGLRRF LMSVAELFVG GV..AVEWSG VF........
atdebs03p   .DAVVV.GSL HRDGGDLSAF LRSMATAHVS GV..DIRWDV AL........
 atnid04p   .GAVAV.GSL RRDDGGLRRF LTSAAEAQVA GV..PVDWAA LC........
atdebs05p   AGACVV.GTL RRDRGGLADF HTALGEAYAQ GV..EVDWSP AF........
atdebs04p   AEVTCV.PTL RREQSGPHEF LRNLLRAHVH GVGADL.... ..........
```

Fig 2s

```
atave02a  PQTH..... .......... .....LLTNL AK......TT T..TWHPHHY
atave05a  PQTH..... .......... .....LLTNL AK......TT T..TWHPHHY
atave04a  PQTH..... .......... .....LLTNL AK......TT T..TWHPHHY
atave08a  PQTH..... .......... .....LLTNL AK......TT T..TWHPHHY
atave03a  PQTH..... .......... .....LLTNL AK......TT T..TWHPHHY
atrap02a  ......IAML HGD.HE.... ..AQAAVGAL AHLYVNG.VS V..EW.SAVL
atrap11a  ......VAML HGD.HE.... ..AQAAVGAL AHLYVNG.VS V..EW.SAVL
atrap08a  ......VAML HGD.HE.... ..AQAAVSAL AHLYVNG.VT V..DW.PALL
atrap12a  ......VAML HGD.HE.... ..IQAAIGAL AHLYVNG.VT V..DW.PALL
atrap05a  ......VAML HTD.HE.... ..AQAAISAL AHLYVNG.VT V..DW.TALL
atrap09a  ......VAML HGD.HE.... ..TQAAIGAL AHLYVNG.VT V..DW.TALL
atfkb03a  ......IPVL HGE.DE.... ..ARSAMTAL ARLHTGG.VA V..DW.PEVI
atfkb07x  ......IPTQ TGTPEE.... ..VQALHTAL ARLHTRG.GV V..DW.PTVL
atfkb08x  ......IALQ NGTADE.... ..VHALHTAL ARLFTRG.AT L..DW.SRIL
atnid01a  ....TTIPTL HREHPEPETL TTAL....AT ..LHTTGHTT T.........
atnid03a  ....TTIPTL HREHPEPETL TTAL....AT ..LHTTGHTT T.........
atnid02a  ....TTIPTL HREHPEPETL TTAL....AT ..LHTTGHTT T.........
atnid00a  ....TTIPTL HRERPEPETL TQAI....AA VGVRTDGIDW A.........
atfkb10a  .......AVL RARTGEES.. ....AALTAV AELHAHG.AP V..DL.AAVL
atrap14a  .......AVL RPRSPEDV.. ....CLMTAI AELHAGG.TA I..DW.AKVL
atmon06a  ....VLVASL AGERPEES.. ....AFVEAM ARLHTAG.VA V..DW.SVLF
atmon08a  ....VLVSSL AGERPEES.. ....AFVEAM ARLHTAG.VA V..DW.SVLF
atmon09a  ....VVTASL HPDRPDDV.. ....AFAHAM ADLHVAG.IS V..DW.SAYF
atepo02a  ....TLLASL RAGREEA... ...AGVLEAL GRLWAAGGS. V..SW.PGVF
atepo03x  ....TLLASL RAGREEA... ...AGVLEAL GRLWAAGGS. V..SW.PGVF
atepo08a  ....VLLPAS RAGRDEA... ...ASALEAL GGFWVVGGS. V..TW.SGVF
atepo00a  ....ALLASS RAGRDEP... ...ATVLEAL GGLWAVGGL. V..SW.AGLF
atepo04a  ....VLVPSL RADRSEC... ...EVVLAAL GAWYAWGGA. L..DW.KGVF
atnid07a  AFAAALRRGR ....PEC... ...ATVLPAA ATAFVQG.AH V..DW.AAPY
atty107a  LLPAIHKPGT APHGPAA... ...PGALRAA AAAYGRG.AR V..DW.AGMH
atsor02a  PCAFL..PTL RKGRDDA... ...EAFTAAL GALHAAG.LT P..DW.SAFF
atsorb1a  PCAFL..PTL RKGRDDA... ...EAFTAAL GALHAAG.LT P..DW.NAFF
atnys09a  .AVTL..PAL RAGRPEE... ...HTLTTAL AGLHVHG.AT L..DW.TGCF
atnys12a  .AL.L..PTL RGDRPEE... ...PALVTAV AAAHAHG.AR V..DW.SGYF
atnys16a  .L.VT..PTL RKDRDEE... ...SALLAGL ARLHVAG.VT V..DW.SAAL
atnys17a  .E.VV..PAL RKGRPEE... ...HTALTAA AQLHVAG.VD I..DW.TAVL
atnys03a  .A.AV..PLL RKDRPEE... ...LSAVTGL ARAHVRG.VT V..RW.AGLF
atnys15a  .V.SV..PVL RKDRDEE... ...PAAVAAL ARLHTAG.VP V..DW.TAFY
atnys07a  .L.AV..PLL RKDRPEE... ...PAALAAL AQLHIAG.AR V..DW.PVLF
atnys08a  .V.TV..PVL RKDRGEE... ...STALTAR AHLHTRG.LI E..DW.QDFF
atnys05a  .A.TV..PAL RKDRDEE... ...TSALTAL AHLHTAG.LR V..DW.AAFF
atnys06a  .V.TV..PVL RRNMPEE... ...RTLLTAL GRLHTTG.TP I..DW.AALL
atnys04a  .G.TI..PLL RRDRPEE... ...QAVLAAL CHLQVLG.VE A..DW.SATF
atnys14a  .V.VV..PAL RRNRDED... ...ETLVGAV ARLHVHG.AG P..RW.DAYF
atnys00a  .TDVV..PAL SKGRPEE... ...TAFAGAL GRLHTLG.VP V..DW.PAFY
atnys10a  .AELV..PML RAGRAEE... ...LAAATAL ARLQVRG.VD V..DW.AAYL
atnys18a  .LVAV..PVL RKERPEE... ...TTVLAAL GTLWAHG.AD V..DW.DAVF
atnys13a  PAVVV..PLQ RRDRAGD... ...LALLEGL ATLHTHG.TG P..SW.PAYF
atave10a  .ARAI..PAL RPDQPEA... ...RSVMTAL AELFVAG.TA V..EW.AGVF
atrif02a  ....V..ATL RKNGAEV... ...PDVLTAL AELHVRG.VG V..DW.TTVL
atmon03a  EPEPVVAAAL RSKHDEG... ...RTLLGAV AALHTDG.QP A..DL.TALF
atave12a  PQENLLIPLL RPDSPEP... ...GTLLTGL ARLHTHGAAA V..NW.PAAL
atrif09a  .AEVTCVAAL RDDRPEV... ...TALITAV AELFVRG.VA V..DW.PALL
atmon00a  ....AAIATC RRGRDEV... ...ATFLRSL AQAYVRG.AD V..DF.TRAY
atty103a  PDPLLTLPLL RRSVPETGDA EHPGGFERAL ATAYAHGV.. ......PLRL
```

Fig 2t

```
          501                                                        550
 atave00x  TQQQSLLLDL VRAHTMAVLN DDGN~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atdebs00p  GSLPE.FAPA EEEDEPAESG VDWNAPPHVL RER~~~~~~~ ~~~~~~~~~~
 atepo06p  .......... RRVPLPTYPW QHERCWIEVE PDARR~~~~~ ~~~~~~~~~~
 atepo07p  .......... RRVPLPTYPW QHERYWIEDS VHGSKPSLRL RQLRNGATDH
 atepo01p  .......... RRVPLPTYPW QRERYWIEAP AKSAAGDRRG VRAGGHPLLG
 atepo05p  .......... RRVPLPTYPW QRERYWVDAP TGGAAGGSRF AHAGSHPLL~
 atsora1x  .......... ESMPLPSTAG ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
 atfkb01p  .......P.H RVLDLPTYPF DHKRYWLQPA PVT~~~~~~~ ~~~~~~~~~~
 atfkb09p  .......P.D RLLDLPTYPF DHKRYWIEAT GAADLTALGL TDTAHP~~~~
 atrap03p  .......TTT RVLDLPTYAF QHQRYWVE.. .GVDRSAAG. ...GHPLLGV
 atrap06p  .......ATT RVLDLPTYAF QHQRYWLR.. .SVDRAAAD. ...GHPLLGT
 atrap04p  .......TTA RVLDLPTYAF QHQRYWVK.. .SVDRAAAD. ...GHPLLGA
 atrap13p  .......TTT RVLDLPTYAF QHQRYWLK.. .SVDRAAAD. ...GHPLLGT
 atrap01p  .......TAA RVLDLPTYAF QHQRYWLK.. .GVDRAAAD. ...GHPLLGT
 atrap07p  .......ATT RVPDLPTYAF QHQRFWAE.. .GADRSVAG. ...GHPLLGV
 atrap10p  .......PAT RVLDLPTYAF QHQRYWAEAG RSADVSAAGL DAVGHPLLGA
 atfkb04x  AASVTAHDTG TAHDLPTYAF HHERYWIEPA TGTDASGLGL D~~~~~~~~~
 atty104p  ...DPALPPG HLTTLPTYPF NHHHYWLDTT PTTPA.TTTQ SPTDAQNPAD
 atty106p  ...DPALPPG HLTTLPTYPF NHHHYWLDTT PTTPA.TTTQ SPTDAWR...
 atty101p  ...DPALPPG HLTTLPTYPF NHHHYWLDTT PTTPA.TTTQ SPTDAWR...
 atty102p  ...DPALPPG HLTTLPTYPF NHHHYWAVTS PAGVG.DAA. ....AGR...
 atty100p  ...DPALPPG HLTTLPTYPF NHHHYWLDTI DGGGGDDATQ EKESGPLTRE
 atnid05b  .G.....PGA RPVELPTYPF QRERYWCHP. GVRGGDPASL GMDGADHPLL
 atty105b  .G.....PDS RRVNPPTYPF QRERYWYHPT SGRRGDITAA GVAEAEHPLL
 atnid06x  EG.....TGA RRVPLPSYAF HRDRFWLPTA AARRPATSSS ~~~~~~~~~~
atdebs01p  LG.....TGA RRVPLPTYPF QRERVWLEPK PVARRSTEVD EV~~~~~~~~
 atmon02p  ....PADPAP RTIDLPTYAF QRRRYWLADT VKRDSGWDPA GS~~~~~~~~
 atmon10p  ....PADPTP RTVDLPTYAF QHQHYWLERS ASASGAVSGE QSA~~~~~~~
 atmon04p  ....PADPTP RTVDLPTYAF QHQHYWLEEP SGLTGDAADL GMVA~~~~~~
 atmon07p  ....PADPTP RTIDLPTYAF QRRSYWL..P VDGVGDVRSA GLRRVE~~~~
 atmon11p  ....PADPTP RTVDLPTYAF QHKHYWVEPP AAVAAVGGGH DPVEA~~~~~
 atmon12p  ....PAVPLP RVVDLPTYAF QRERFWLEGR RGLAGDPAGL GL~~~~~~~~
 atmon05b  ....PADPAP RTVDLPTYAF QRQDFWPAPA GGRSGDPAGL GLAASGHP~~
 atmon01p  ....PATGT. STVELPTYAF QRRRYWLEAP TG.TQDAAGL GL~~~~~~~~
atdebs02p  ........GRA GLVDLPGYPF QGKRFWLLPD RTTPRDEL.D GWF~~~~~~~
atdebs06p  .......PAA PPVALPNYPF EPQRYWLAPE VS...DQLAD SRYRVD~~~~
 atave01p  THHHTHPHPH THLDLPTYPF QHQHYWLESS QPGAGSGSG~ ~~~~~~~~~~
 atave07p  THHHTHPHNH .HLDLPTYPF QRQHYWLD.A PTGAGDV~~~ ~~~~~~~~~~
 atave06p  TQTHPHPNPH THLDLPTYPF QHQHYWLQPP TTTTDLTTTG LTPTHHPL~~
 atave09p  TQTHPHPHNH .HLDLPTYPF QHQHYWLQ~~ ~~~~~~~~~~ ~~~~~~~~~~
 atnys01p  E.....GTGA ARVDLPTYAF QRERYW.NTR TAADRTPADA PMDAEFWA~~
 atnys11p  E.....GTGA SRIDLPTYAF QHEHLW.AVP PAPEAVAAAD PDDAAFWTAV
 atrif05p  .......PPA .RVELPTYAF DHQHFW..LS PPAVA.DAPA LGLAGADHPL
 atrif07p  .......PPA .RVDLPTYAF DHQHFW..LR PAAQA.DAVS LGQAAAEHPL
 atrif08p  .......PPA .RVDLPTYAF DHQHYW..LR YVETATDAA~ ~~~~~~~~~~
 atrif10p  .......PPA .RADLPTYAF DHEHYW..LR AADTASDAVS LGLAGADHPL
 atrif03p  .......PAA GWVDLPTYAF DRRHFW..LH EAETAEAAEG M~~~~~~~~~
 atrif06p  .......PRT GRVDLPKYAF DHRHYW..LR PAESATDAAS LGQGAADHPL
 atrif04p  .......PPS RRVELPTYAF DHQHYW..LQ MGGSATDAV~ ~~~~~~~~~~
 atrif01p  .......PAA GWVDLPTYAF EHRHYW..LE PAEPASAGDP LLGT~~~~~~
 atnys02p  A.....RSAY QPVDLPTYPF QRQDFWPEAR PATPAAGADA SD~~~~~~~~
 atfkb02p  P.....FGEL RGVPLPTYPF RRDRYWVDAE PAGTSGHP~~ ~~~~~~~~~~
 atave11p  GSVGRGVAGG CGVELPTYAF ERERFWLDVE GAPRGSGVSG QW~~~~~~~~
atdebs03p  .......PGA APFALPTYPF QRKRYWLQPA APAAASDELA YRV~~~~~~~
 atnid04p  .......PRA GWVDLPTYAF QRERYWVAPA EPGPAAGAGS AAATGPAAA~
atdebs05p  .......ADA RPVELPVYPF QRQRYWLPIP TGGRARDEDD DWR~~~~~~~
atdebs04p  ...RPAVAGG RPAELPTYPF EHQRFWPRPH RPADVSALGV R~~~~~~~~~
```

Fig 2u

```
atave02a  THHDNQPHTH THLDLPTYPF QHHHYWLE.. STQPGAGNV~ ~~~~~~~~~~
atave05a  THHHNQPHTH THLDLPTYPF QHHHYWLELP SAQTSPGQRR SRRSAPD~~~
atave04a  THHHNQPHTH THLDLPTYPF QHQHYWLE.. STQPGAGSGS GSGSGRAG~~
atave08a  THHHNQPHTH THLDLPTYPF QHHHYWLE.. STQPGAGNVS AA~~~~~~~~
atave03a  THHHNQPHTH THLDLPTYPF QHHHYWLQ.. ..PPGKPSDP SP~~~~~~~~
atrap02a  GDVPVTRV.. ..LDLPTYAF QHQRYWLE.. .GTDRATAG. ...GHPLLGS
atrap11a  GDVPVTRV.. ..LDLPTYAF QHQRYWLE.. .GTDRATAG. ...GHPLLGS
atrap08a  GDAPATRV.. ..LDLPTYAF QHQRYWLE.. .GTDRMAAG. ...GHPLLGE
atrap12a  GDAPATRV.. ..LDLPTYAF QHQRYWLE.. .GTDRATAG. ...GHPLLGS
atrap05a  GDAPATRV.. ..LDLPTYAF QHQRYWLE.. .GADRAAAG. ...GHPLLGP
atrap09a  GDVPVTRV.. ..LDLPTYAF QQQRYWAEVG RSADVSGAGL DAVGHPLLGA
atfkb03a  GAAP.TDL.. ..PHLPTYPF ERTRYWLGSR AAGDA~~~~~ ~~~~~~~~~~
atfkb07x  .GSDRAPV.. ...ALPTYPF QHKDYWLRAT AQVDVTGAGQ EKVAHPLL~~
atfkb08x  GGASRHDP.. ...DVPSYAF QRRPYWIE.S APPATADSG. ....HPVLGT
atnid01a  ..LHTTSPQT HHLDLPTYPF QRDRYWM.EP VRVAQVSGQP GADRLRYRVV
atnid03a  ..LHTTSPQS HHLDLPTYPF QRDRYWM.AV PPRAAVGDLA ~~~~~~~~~~
atnid02a  ..PHPSHIPA QRVSLPAYPF QRRAYWM..P NSAAHIGRSD AEAATRLGLA
atnid00a  ..VLCGASRP RRVELPTYAF QRRTHWAPGL TPNHAPADRP AAEPQRAMAV
atfkb10a  A.......GG RPVDLPVYPF QHRSYWLAPA VGGGSPTAVP D~~~~~~~~~
atrap14a  S.......GG RAVDLPVYPF QHQSYWLAPA ..APDATAVA PVVEEEGGEY
atmon06a  AGDRVPGL.. ..VELPTYAF QRERFWLSG. RSGGGDAATL GLVAAG~~~~
atmon08a  AGDRVPGL.. ..VELPTYAF QRERFWLSG. RSGGGDAATL GLVAAGHPL~
atmon09a  PDDPAPRT.. ..VDLPTYAF QGRRFWLADI AAPEAVSSTD GEEA~~~~~~
atepo02a  ......PTAG RRVPLPTYPW QRQRYWIEAP AE~~~~~~~~ ~~~~~~~~~~
atepo03x  ......PTAG RRVPLPTYPW QRQRYWPDIE PDSRR.HAAA DPTQGWFY~~
atepo08a  ......PSGG RRVPLPTYPW QRERYWIEAP VDREA.DGTG ~~~~~~~~~~
atepo00a  ......PSGG RRVPLPTYPW QRERYWIDTK ADDAA.RGDR RAPGAGHDEV
atepo04a  ......PDGA RRVALPMYPW QRERHWMDLT PRSAA.PAGI AGRWPLAGVG
atnid07a  ...EG..AGA RRVDLPTYPF QHTRYWL~~~ ~~~~~~~~~~ ~~~~~~~~~~
atty107a  A..DGPEGPA RRVELPVHAF RHRRYWLAPG RAA~~~~~~~ ~~~~~~~~~~
atsor02a  A.....PFAP R~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atsorb1a  A.....PFAP CKVPLPTYTF ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
atnys09a  AGT.....GA RRTDLPTYAF QRRRYWPKAL QSGTA.DLRS VGLGAA~~~~
atnys12a  ADH.....GA RRTTLPTYAF QRERYWPDTT AATSA.HTPG SALDAEFW~~
atnys16a  TGT.....GA RGTDLPTYAF QRERYWPE.. LAAEP.AG.. GGADAADA~~
atnys17a  AGT.....GG RRIALPTYAF QRERYWPS.. LAAQA.PGDA GG~~~~~~~~
atnys03a  DGT.....GA RRADLPTYPF QHQRFWPT.. AAR.A.AQDV TAAGLGAADH
atnys15a  AGT.....GA HRTDLPTYAF QYERYWPK.. ATY.R.PADA TGL~~~~~~~
atnys07a  AGV.....GA GRVELPTYAF QRGWFWPV.. GRVGV.GGDV ~~~~~~~~~~
atnys08a  AGV.....GA GRVELPTYAF QRGWFWPV.. GRVGV.GGDV GAVGLGSAGH
atnys05a  AGS.....GA TRVDLPTYAF QHATYWPT.. GTLPT..AHA AAVGL~~~~~
atnys06a  APT.....GA RPVDLPTYAF QHRPFWPS.. GPRDT..ADA AAVGIAGASH
atnys04a  RGL.....DP VRVDLPTYAF QHRWFWPA.. ARPAR.PDDV RAAGLGAA~~
atnys14a  AGR.....GA QWLDLPTYPF QRGRFWPE.. SLPGA.ASAA PAAGQPA~~~
atnys00a  AGT.....GA RRVELPTYAF QHVRHWPT.. PPRPN.GAGP GALGHPLLG~
atnys10a  AGT.....GA RRTDLPTYAF QHAYYWPQ.. LPTPA.AALA AADPADQQLW
atnys18a  AGT..RTPQA DPVELPTYAF QRARYWPTLG ARHGD.PADL G~~~~~~~~~
atnys13a  EAT.....GG HRTDLPTYAF QRERYWPELG APVAT.APQD PAAW~~~~~~
atave10a  EGTAREVGDG CGVELPTYAF ERERFWLDVE EGSAG.GSGV SGMWGGPLWE
atrif02a  ....DEPATA VGTVLPTYAF QHQRFWVDVD ET~~~~~~~~ ~~~~~~~~~~
atmon03a  ......PADA GQVPLPTYRF QRRRYWRVAP DAAAP.ARAA GLQ~~~~~~~
atave12a  PERDR....A RHLDLPTYAF DHHRYWVDTS AGHPG.DLSA AGLGT~~~~~
atrif09a  PPVTGF.... ..VDLPKYAF DQQHYWLQPA AQATD.AASL GQV~~~~~~~
atmon00a  GAT.....AT RRFPLPTYPF QRERHWPAAA GVGQQ.PETP ELP~~~~~~~
atty103a  APAPDAASLA VAAELPTYAF QRTHYWLDAP AAPAALPAGL DDAGHPLLSA
                    ****                                  LPTY motif
```

Fig2v

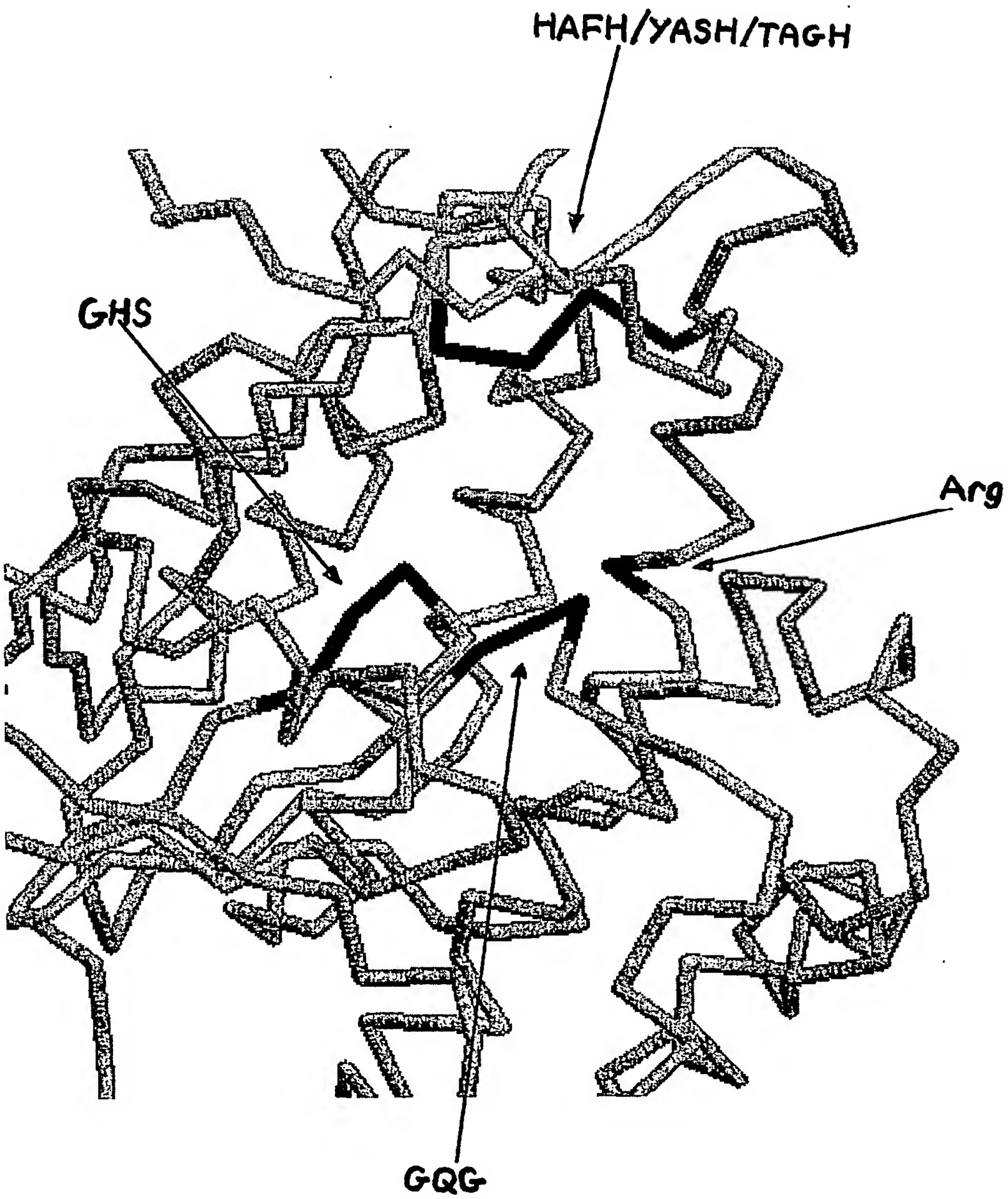


Fig3

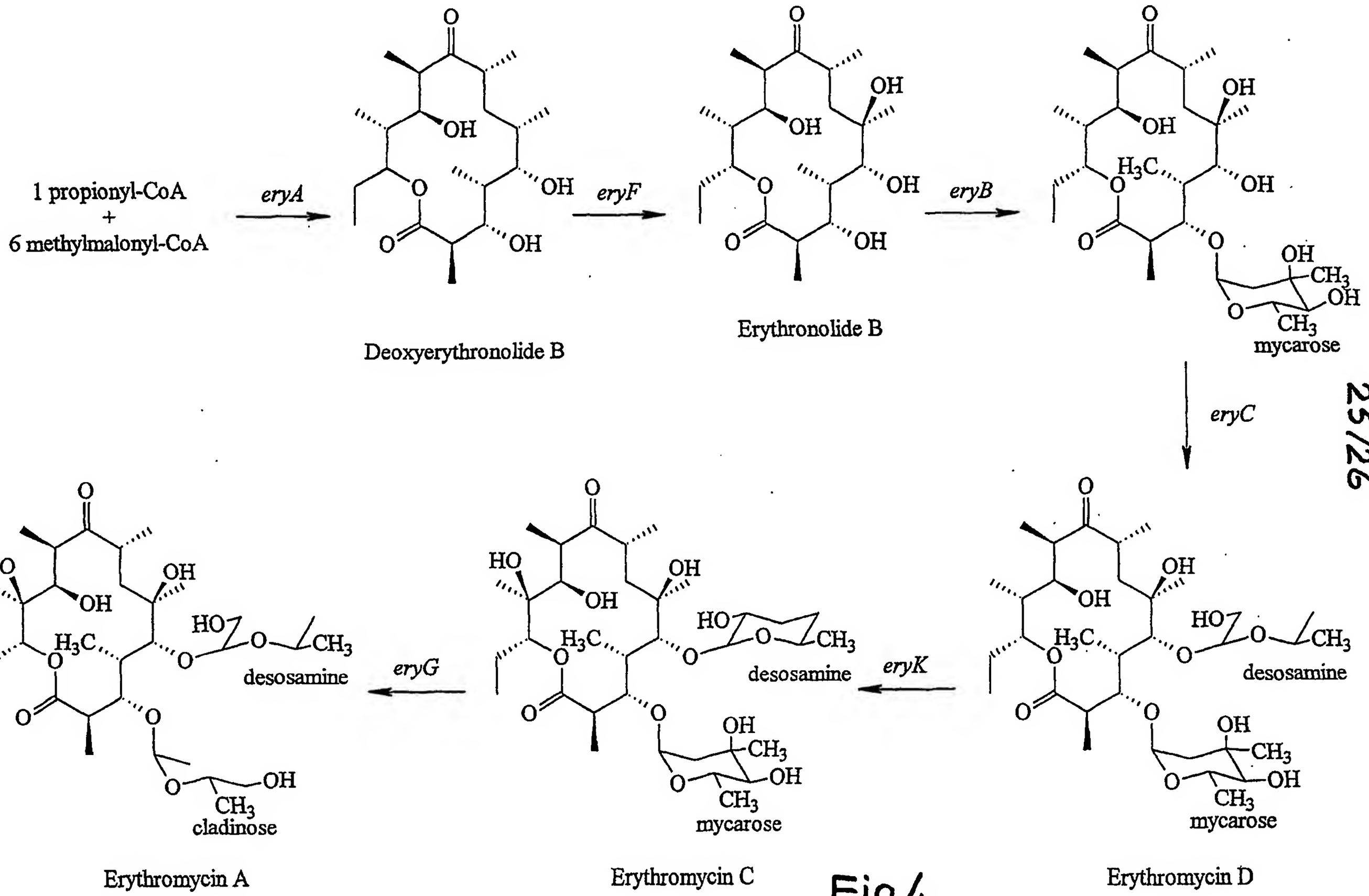

1 propionyl-CoA
+
6 methylmalonyl-CoA
eryA
Deoxyerythronolide B
eryF
Erythronolide B
eryB
mycarose
eryC
desosamine
cladinose
eryG
Erythromycin A
desosamine
mycarose
eryK
Erythromycin C
desosamine
mycarose
Erythromycin D

Fig 4

```
15500  GGTGGTGCCGACGCCATGGCCCGTGAGCGCTCACAGCGCTTCCGCGCTGCGCGCGCAGGC  15559
15560  CGGTCGCCTGCGGACGCACCTCGCCGCCCACCGCCCCACCCCCGACGCCGCGCGGGTCGG  15619
15620  CCACGCGCTCGCCACCACCCGTGCGCCCCTCGCCCACCGCGCGGTCCTGCTCGGCGGCGA  15679
15680  CACCGCCGAACTGCTGGGCTCCCTGGACGCGCTGGCCGAGGGCGCGGAGACCGCGTCCAT  15739
15740  CGTGCGCGGCGAGGCGTACACCGAGGGCAGGACGGCCTTCCTCTTCAGTGGGCAGGGAGC  15799
15800  GCAACGCCTCGGCATGGGGCGGGAGTTGTATGCCGTGTTCCCCGTCTTCGCCGACGCTCT  15859
15860  CGACGAGGCGTTCGCCGCCCTGGACGTACATCTGGACCGCCCACTGCGCGAGATCGTCTT  15919
15920  GGGCGAGACCGACTCGGGTGGGAACGTCTCGGGTGAGAATGTCATCGGCGAGGGTGCCGA  15979
15980  CCATCAGGCACTCCTCGACCAGACCGCCTACACCCAGCCCGCGCTCTTCGCGATCGAGAC  16039
16040  GAGCCTGTACCGGCTGGCAGCCTCCTTCGGCCTGAAGCCGGACTACGTCCTCGGCCACTC  16099
16100  GGTCGGCGAGATCGCCGCCGCGCACGTCGCCGGTGTCCTCTCGTTGCCGGACGCGAGCGC  16159
16160  TCTGGTGGCCACGCGGGGACGGCTCATGCAGGCGGTTCGCGCGCCCGGCGCGATGGCCGC  16219
16220  GTGGCAGGCCACGGCGGACGAGGCGGCCGAACAGCTCGCCGGGCACGAGCGGCACGTCAC  16279
16280  CGTGGCCGCCGTCAACGGCCCCGACTCCGTGGTCGTCTCCGGCGACCGCGCCACCGTCGA  16339
16340  CGAACTGACCGCCGCCTGGCGGGGACGCGGCCGCAAGGCCCACCACCTGAAGGTCAGCCA  16399
16400  CGCCTTCCACTCCCCGCACATGGACCCCATCCTCGACGAGCTGCGCGCGGTCGCCGCCGG  16459
16460  CCTGACCTTCCACGAGCCGGTCATTCCCGTCGTCTCCAACGTCACCGGTGAACTGGTGAC  16519
16520  CGCGACCGCGACCGGGAGCGGCGCCGGGCAGGCCGACCCCGAGTACTGGGCGCGGCATGC  16579
16580  GCGCGAGCCCGTGCGGTTCCTGTCCGGGGTGCGGGGGCTGTGCGAGCGCGGGGTGACCAC  16639
16640  GTTCGTCGAGCTCGGCCCGGACGCACCGCTGTCCGCGATGGCCCGCGACTGCTTCCCCGC  16699
16700  CCCCGCGGACCGGAGCCGTCCGCGCCCCGCCGCCATCGCCACATGCCGCCGCGGGCGCGA  16759
16760  CGAGGTGGCCACGTTCCTGAGGTCGCTGGCCCAGGCGTACGTCCGCGGCGCCGATGTCGA  16819
16820  CTTCACCCGGGCCTACGGCGCCACCGCCACGCGCCGCTTCCCCCTCCCCACGTATCCCTT  16879
16880  CCAGCGCGAGCGCCATTGGCCTGCCGCTGCCGGGGTGGGGCAGCAGCCGGAGACCCCGGA  16939
16940  ACTTCCGGAATCCTCGGAGTCCTCGGAGCAGGCAGGGCATGAGCGGGAGGAGGGGGCGCG  16999
17000  CGCGTGGGGCGGGCCTGAAGGGCGGCTTGCCGGGCTCTCCGTGAACGACCAGGAGCGGGT  17059
17060  CCTCCTCGGCCTGGTCACCAAGCACGTGGCCGTCGTGCTCGGGGACGCCTCGGGCACGGT  17119
17120  ACAAGCCGCCCGCACCTTCAAGCAGTTGGGCTTCGACTCGATGGCCGCCGCCGAGCTGAG  17179
17180  CGAACGGCTCGGCACGGAGACGGGCCTGCCGTTGCCCGCCACCCTCACCTTCGACTACCC  17239
17240  GACCCCTCTGGCCGTCGCCGCGCACCTGCGCGCGGAGCTCACCGGTACGCCCGCCCCGGC  17299
17300  CGGCTCCGCGCCCGCCACGGGCGCCCTCGGCGCGGGTGACCTCGGCACGGACGAGGACCC  17359
17360  GGTCGCCATCG  17370
```

Fig 5